MATHEMATISCHE ZEITSCHRIFT

71. BAND

MATHEMATISCHE ZEITSCHRIFT

UNTER STÄNDIGER MITWIRKUNG VON

E. KAMKE
TÜBINGEN

R. NEVANLINNA
HELSINKI

E. SCHMIDT
BERLIN

F. K. SCHMIDT
HEIDELBERG

HERAUSGEGEBEN VON

H. WIELANDT
TÜBINGEN

WISSENSCHAFTLICHER BEIRAT

W. BLASCHKE L. FEJÉR A. E. INGHAM H. KNESER
W. MAGNUS O. PERRON G. PICKERT

71. BAND

BERLIN · GÖTTINGEN · HEIDELBERG
SPRINGER-VERLAG
1959

Unveränderter Nachdruck 1973

Springer-Verlag Berlin Heidelberg New York

Inhalt des 71. Bandes

Seite

AHRENS, J., Begründung der absoluten Geometrie des Raumes aus dem Spiegelungsbegriff 154

BAER, R., Kriterien für die Abgeschlossenheit endlicher Gruppen 325

BAER, R., Direkte Produkte von Gruppen teilerfremder Ordnung 454

BAIER, O., Über die Konstruktion von Parallelen mit Hilfe des Lineals und ortsfester rechter Winkel 94

BAKER, I. N., Fixpoints and Iterates of Entire Functions 146

BARNER, M., und W. DEGEN, Über sich schließende Figuren asymptotisch transformierter Kurven und W-transformierter Flächen 361

COHN, P. M., On the free Product of Associative Rings 380

ERWE, F., Abschätzungen für beschränkte Funktionensysteme 414

FOSTER, A. L., An Existence Theorem for Functionally Complete Universal Algebras 69

GAIER, D., Über ein Extremalproblem der konformen Abbildung 83

GASCHÜTZ, W., Fixkörper von p-Automorphismengruppen rein-transzendenter Körpererweiterungen von p-Charakteristik 466

GERSTENHABER, M., Note on a Theorem of Wielandt 141

GÜNZLER, H., Fastperiodische Lösungen linearer hyperbolischer Differentialgleichungen 223

HADWIGER, H., Normale Körper im euklidischen Raum und ihre topologischen und metrischen Eigenschaften 124

HARTMAN, P., On solutions of $\Delta V + V = 0$ in an exterior region 251

HENRIKSEN, M., Multiplicative summability methods and the Stone-Cech compactification 427

HOFMANN, K. H., Topologische distributive Doppelloops 36

KELLERHALS, G., Randwertprobleme bedingter Wahrscheinlichkeiten 1

KEOGH, F. R., and G. M. PETERSEN, A strengthened form of a theorem of WIENER 31

KOECHER, M., und W. ROELCKE, Diskontinuierliche und diskrete Gruppen von Isometrien metrischer Räume 258

KOPPELMAN, W., and J. D. PINCUS, Spectral Representations for Finite Hilbert Transformations 399

LEICHTWEISS, K., Selbstadjungierte Banach-Räume 335

MACROBERT, T. M., Infinite Series of E-Functions 143

NICKEL, K., Fehlerabschätzungen bei parabolischen Differentialgleichungen 268

OSTROM, T. G., and A. WAGNER, On projective and affine planes with transitive collineation groups 186

Seite

PETERSSON, H., Über Darstellungsanzahlen von Primzahlen durch Quadratsummen 289

PICKERT, G., Eine Kennzeichnung desarguesscher Ebenen 99

REMBS, E., Über die Verbiegbarkeit der Rinnen 89

SALZMANN, H., Topologische Struktur zweidimensionaler projektiver Ebenen . . . 408

TACHIKAWA, H., On Rings for which Every Indecomposable Right Module has a Unique Maximal Submodule 200

TAUSSKY, O., A weak property L for pairs of matrices 463

VIDAV, I., Zur Axiomatik des Hilbertschen Raumes 458

WAGNER, A., On perspectivities of finite projective planes 113

WALTER, W., Über die Differentialgleichung $u_{xy}=f(x, y, u, u_x, u_y)$ 308

WALTER, W., Über die Differentialgleichung $u_{xy}=f(x, y, u, u_x, u_y)$. II. Existenzsätze für das charakteristische Anfangswertproblem 436

WEVER, F., Über stationäre Null-Eins-Prozesse mit beschränkten Lücken 283

WIELANDT, H., Zum Satz von Sylow. II 461

ZELLER, K., Saturation bei äquivalenten Summierungsverfahren 109

KELLERHALS, G.
Math. Zeitschr. 71, 1—30 (1959)

Randwertprobleme bedingter Wahrscheinlichkeiten

Von
GERHARD KELLERHALS

Problemstellung

Das I. Kapitel dieser Arbeit beschäftigt sich mit Wahrscheinlichkeiten (Wn) von stochastisch abhängigen Ereignissen, die den Punkten einer eindimensionalen Punktmenge ($x=0, 1, \ldots, n$) zugeordnet sind. Jeder dieser Punkte kann das Merkmal 0 oder 1 annehmen. Die zugehörigen Ereignisse seien A_0^x und A_1^x; für ihre Wn gilt $P(A_0^x)+P(A_1^x)=1$. Die eindimensionale Anordnung ist wesentlich für die Art der gegenseitigen Abhängigkeit der Ereignisse A_i^x, für die gelte: Unter der Annahme, daß irgendwelche der Ereignisse $A_{i_\nu}^{x_\nu}$ links und rechts von A_i^x $(0<x<n)$ bekannt sind, sei A_i^x nur von dem nächstgelegenen linken und dem nächstgelegenen rechten dieser Ereignisse abhängig. Besitzen die Ereignisse $A_{i_0}^0, A_{i_1}^1, \ldots, A_{i_n}^n$ diese Eigenschaft, so soll der zugehörige Wahrscheinlichkeitsraum ein diskreter, alternativer Nachbarprozeß (NP) heißen (engl. Nearest neighbour system)[1]. Trotz der Ausnahmestellung der Punkte $x=0$ und n in der Definition zeigt sich, daß keiner der Punkte $x=0, 1, \ldots, n$ vor dem andern ausgezeichnet ist. Die Punkte n und 0 sind demnach ebenso benachbart wie x und $x+1$, und dies entspricht in den Anwendungen einer ringförmigen Anordnung der merkmaltragenden Punkte.

Erst wenn man zusätzlich fordert, daß A_i^0 von A_λ^n nicht abhängt, wenn $A_\varkappa^x$ $(0<x<n)$ bekannt ist, nehmen die Punkte 0 und n eine Randstellung ein, was anschaulich durch Abtragen der Punkte $x=0, 1, \ldots, n$ auf einem Geradenstück oder einer offenen Kurve zum Ausdruck gebracht werden kann. In diesem Fall soll der NP offen heißen.

Die Art der Abhängigkeit legt es nahe, bei den NP die bedingten Wn $P(A_\varkappa^x | A_i^{x-1} A_\lambda^{x+1})$ für $i, \varkappa, \lambda = 0, 1$ und $x=0, 1, \ldots, n$, wobei $n+1$ durch 0 und -1 durch n zu ersetzen ist, vorzugeben. Es zeigt sich, daß durch diese Größen ein NP bestimmt ist, wenn dieselben noch gewisse Verträglichkeitsbedingungen erfüllen. Bei den offenen NP arten die Verkettungs-Wn für $x=0$ und n in die bedingten Rand-Wn $P(A_\varkappa^0 | A_\lambda^1)$ und $P(A_\varkappa^n | A_i^{n-1})$ aus.

Bei der Berechnung der Wn $P(A_i^x)$, $P(A_i^x A_\varkappa^{x-1}), \ldots$, der Aufstellung der Verträglichkeitsbedingungen, der Untersuchung des asymptotischen Verhaltens der homogenen offenen NP $(P(A_\varkappa^x | A_i^{x-1} A_\lambda^{x+1}) = \text{const}$ bezüglich $x)$ und allen weiteren Ausführungen wird neben den Wn das sog. Wahrscheinlichkeitsverhältnis $Q(A) = P(A) : P(\bar{A})$ verwendet.

Sehr wichtig ist der Zusammenhang mit den alternativen Markoff-Ketten (MK), die in der Klasse der NP echt enthalten sind. Jedoch wird gezeigt werden, daß jeder offene NP auch eine MK ist. Dies hängt damit zusammen,

[1]) Siehe M. BARTLETT [1], S. 35.

daß die Ereignisse A_i^x einer MK nicht nur von vergangenen, sondern auch von zukünftigen Ereignissen stochastisch abhängen, und daß die zeitliche Umkehrung eines Markoff-Prozesses wieder ein solcher ist.

Die für die numerische Behandlung und für allgemeine Aussagen wichtigsten Klassen der homogenen MK $(P(A_\varkappa^x | A_i^{x-1}) = \text{const bezüglich } x)$ und der homogenen NP sind jedoch nicht äquivalent; vielmehr sind die ersteren in der Klasse der letzteren echt enthalten[2]). (Bei den homogenen NP sind vier Größen frei wählbar, bei den homogenen MK nur drei.)

Zum Unterschied zwischen den MK und den offenen NP bezüglich der Anwendungen sei folgendes gesagt: Bei den MK wird eine zeitliche Aufeinanderfolge der Ereignisse A_i^x oder wenigstens die Auszeichnung einer Richtung angenommen, was auch durch Vorgabe der Wn $P(A_i^0)$, $P(A_\varkappa^x | A_i^{x-1})$ zum Ausdruck kommt. Außerdem ist der Begriff homogen bei MK gegenüber einer Umkehrung der Fortschreitungsrichtung nicht invariant. Demgegenüber soll bei den NP im allgemeinen keine Richtung bevorzugt werden, also insbesondere keine zeitliche Anordnung der Ereignisse A_i^x vorausgesetzt sein. Eine ausgezeichnete Richtung ist auch in vielen Anwendungsbeispielen nicht vorhanden oder ohne zusätzliche Angaben nicht erkennbar. So kann etwa bei einer Baumreihe einer Zuchtpflanzung, deren Bäume auf Befall von einer ansteckenden Krankheit untersucht werden, nicht ohne weiteres von einer bevorzugten Ausbreitungsrichtung gesprochen werden.

Aus diesen Gründen wird man bei einer zeitlichen Aufeinanderfolge von abhängigen Ereignissen an eine Anwendung von MK denken; dagegen sind die NP auf räumlich angeordnete Ereignisse zugeschnitten.

Auf die Möglichkeiten einer Anwendung der NP in Biologie, Physik usw. soll hier nicht eingegangen werden. Jedoch sei auf die Arbeiten von E. MONTROLL [4], [5] und die dort aufgeführte Literatur hingewiesen.

Wie die MK, so lassen sich auch die NP in mehrfacher Weise verallgemeinern. Einmal kann der Wertevorrat der Merkmale jedes Punktes von $\{0, 1\}$ auf $\{1, 2, \ldots, N\}$, $\{1, 2, 3, \ldots\}$ oder auf ein Kontinuum erweitert werden. Zum andern kann die diskrete Ortskoordinate $(x = 0, 1, \ldots, n)$ durch eine kontinuierliche $(0 \leqq x \leqq d)$ ersetzt werden. Und schließlich lassen sich beide Möglichkeiten noch kombinieren.

Zu der Erweiterung auf nichtalternative Merkmale sei nur gesagt, daß sich die für alternative Merkmale abgeleiteten Formeln in gewissem Sinn verallgemeinern lassen, aber auf eine Aufnahme in diese Arbeit verzichtet wurde.

Dagegen werden im II. Kapitel NP mit stetiger Ortskoordinate und alternativen Merkmalen behandelt. Den im I. Kapitel abgeleiteten Rekursionsformeln entsprechen hier Funktionalgleichungen, die bei gewissen Differenzierbarkeitsvoraussetzungen durch Differentialgleichungen vom Riccati-Typ ersetzt werden können, welche in Analogie zu den vor- und rückwärtsgenommenen Kolmogoroffschen Differentialgleichungen bei MK stehen.

Für die Anregung zu dieser Arbeit und die Unterstützung während ihrer Durchführung sei Herrn Professor Dr. GÜNTHER SCHULZ herzlich gedankt.

[2]) Siehe auch die Übersicht am Schluß von § 8.

I. Kapitel

Diskrete alternative Nachbarprozesse

§ 1. Definition des diskreten, alternativen Nachbarprozesses

Ein Ereignisraum bestehe aus der Gesamtheit der Punkte (Elementarereignisse) $(\iota_0, \iota_1, \ldots, \iota_x, \ldots, \iota_n)$ mit $\iota_x = 0, 1$ und $n > 2$. Das Ereignis, an der Stelle x eine 1 bzw. 0 zu erhalten, sei mit A_1^x bzw. A_0^x bezeichnet. Statt A_ι^x wird bei den folgenden Aussagen, die für $\iota = 0$ und $\iota = 1$ gelten, im allgemeinen nur A^x geschrieben. Jedem der Elementarereignisse $(\iota_0, \iota_1, \ldots, \iota_n)$ oder $A_{\iota_0}^0 A_{\iota_1}^1 \ldots A_{\iota_n}^n$ sei eine Wahrscheinlichkeit (W) $P(A_{\iota_0}^0 A_{\iota_1}^1 \ldots A_{\iota_n}^n) > 0$ zugeordnet, so daß

$$\sum_{\iota_0=0}^{1} \sum_{\iota_1=0}^{1} \cdots \sum_{\iota_n=0}^{1} P(A_{\iota_0}^0 A_{\iota_1}^1 \cdots A_{\iota_n}^n) = 1$$

ist.

Aus der Gesamtheit der dadurch erhaltenen Wahrscheinlichkeitsräume oder stochastischen Prozesse werden im folgenden die sog. Nachbarprozesse (NP) untersucht.

Definition 1. *Die Ereignisse $A^0, A^1, \ldots, A^n$ bilden einen diskreten, alternativen NP, wenn für die bedingten Wn*

$$P(A^{x_\nu} | A^{x_1} \ldots A^{x_{\nu-1}} A^{x_{\nu+1}} \ldots A^{x_m}) = P(A^{x_\nu} | A^{x_{\nu-1}} A^{x_{\nu+1}})$$

für alle ganzzahligen $x_0, x_1, \ldots, x_m$ mit

$$0 \leqq x_0 < \cdots < x_\nu < \cdots < x_m \leqq n$$

gilt.

Ist also das Eintreffen gewisser Ereignisse links *und* rechts von A^x $(x \neq 0, n)$ bekannt, so ist A^x nur von dem nächstgelegenen linken und nächstgelegenen rechten dieser Ereignisse abhängig. Diese Erklärung steht bei Bartlett [1], S. 35 als Definition des nearest neighbour system. Dagegen wird dort die Aussage von Definition 1 als Satz für MK bewiesen.

Für das Folgende sollen die x_ν, wie eben, stets ganze Zahlen mit der in der Definition angegebenen Einschränkung sein.

Eine Aussage für den Fall, daß nur rechts oder nur links von x_ν gewisse Ereignisse bekannt sind, macht der folgende

Satz 1. *Für NP gilt bei jeder Wahl der $x_0, x_1, \ldots, x_\rho$*

$$P(A^{x_0} | A^{x_1} A^{x_2} \ldots A^{x_\rho}) = P(A^{x_0} | A^{x_1} A^{x_\rho})$$

und

$$P(A^{x_\rho} | A^{x_0} A^{x_1} \ldots A^{x_{\rho-1}}) = P(A^{x_\rho} | A^{x_0} A^{x_{\rho-1}})$$

Das Ergebnis dieses Satzes ist in der Deutung, die der Definition folgte, enthalten, wenn man sich die Stellen $x = 0, 1, 2, \ldots, n$, an denen die Ereignisse A^x beobachtet werden, auf einer geschlossenen, doppelpunktfreien Kurve (etwa einem Kreise) so aufgetragen denkt, daß x zwischen[3]) $x-1$ und $x+1$

[3]) Damit der Begriff „zwischen a und b" hier eindeutig wird, denke man sich den Kreis an der Stelle $a - \varepsilon$ aufgeschnitten. Die Ausdrücke „zwischen a und b" und „zwischen b und a" sind also hier zu unterscheiden.

und 0 zwischen n und 1 zu liegen kommt. A^x ist also von A^y unabhängig, wenn das Eintreffen von je einem Ereignis, das den Stellen zwischen x und y und zwischen y und x zugeordnet ist, bekannt ist. Und dies gilt für beliebige Stellen x und y. Es ist also keine der Stellen $0, 1, \ldots, n$ vor der anderen ausgezeichnet.

Beweis zu Satz 1. Es gilt allgemein

$$P(A^{x_0}A^{x_1}\ldots A^{x_{p-1}}\,|\,A^{x_0}A^{x_1}A^{x_p})$$
$$= P(A^{x_0}\,|\,A^{x_0}A^{x_1}A^{x_p})\cdot P(A^{x_1}\,|\,A^{x_0}A^{x_1}A^{x_1}A^{x_p})\ldots P(A^{x_{p-1}}\,|\,A^{x_0}A^{x_1}\ldots A^{x_{p-1}}A^{x_p}).$$

Für NP kann aber auf der rechten Seite überall A^{x_0} gestrichen werden. Danach läßt sich dieselbe zusammenfassen in

$$P(A^{x_1}A^{x_2}\ldots A^{x_{p-1}}\,|\,A^{x_1}A^{x_p}).$$

Andererseits gilt allgemein $(P(A^{x_0}A^{x_1}A^{x_2}\ldots A^{x_{p-1}}\,|\,A^{x_1}A^{x_p})=)$

$$P(A^{x_0}\,|\,A^{x_1}A^{x_p})\cdot P(A^{x_1}A^{x_2}\ldots A^{x_{p-1}}\,|\,A^{x_0}A^{x_1}A^{x_p})$$
$$= P(A^{x_0}A^{x_2}\ldots A^{x_{p-1}}\,|\,A^{x_1}A^{x_p})\cdot P(A^{x_0}\,|\,A^{x_1}A^{x_2}\ldots A^{x_p}).$$

Nach dem eben Bewiesenen sind aber der zweite Faktor links und der erste Faktor rechts dieser Gleichung einander gleich und damit auch, wie behauptet, die beiden andern. Genau so läßt sich die zweite Aussage herleiten.

In Analogie zur Definition der Unabhängigkeit von Ereignissen A^x durch Gültigkeit der Produktregel

$$P(A^{x_0}A^{x_1}\ldots A^{x_p}A^{x_{p+1}}\ldots A^{x_m}) = P(A^{x_0}\ldots A^{x_p})\,P(A^{x_{p+1}}\ldots A^{x_m})$$

für beliebige Produkte von Ereignissen A^x sei noch eine zweite Definition für NP angegeben.

Satz 2. *Die Ereignisse $A^0, A^1, \ldots, A^n$ bilden genau dann einen NP, wenn*

$$P(A^{x_0}A^{x_1}\ldots A^{x_\varkappa}\ldots A^{x_\lambda}\ldots A^{x_m}) = \frac{P(A^{x_\varkappa}\ldots A^{x_\lambda})\,P(A^{x_\lambda}\ldots A^{x_m}A^{x_0}A^{x_1}\ldots A^{x_\varkappa})}{P(A^{x_\varkappa}A^{x_\lambda})}$$

für beliebige $x_0, x_1, \ldots, x_m$ und $\varkappa<\lambda$ gilt.

Beweis. 1) notwendig: Es ist

$$P(A^{x_0}\ldots A^{x_\varkappa}\ldots A^{x_\lambda}\ldots A^{x_m})$$
$$= P(A^{x_\varkappa}\ldots A^{x_\lambda})\,P(A^{x_{\lambda+1}}\ldots A^{x_m}A^{x_0}\ldots A^{x_{\varkappa-1}}\,|\,A^{x_\varkappa}\ldots A^{x_\lambda}).$$

Ähnlich wie beim 1. Teil des vorigen Beweises zeigt man leicht, daß für NP

$$P(A^{x_{\lambda+1}}\ldots A^{x_m}A^{x_0}\ldots A^{x_{\varkappa-1}}\,|\,A^{x_\varkappa}\ldots A^{x_\lambda})$$
$$= P(A^{x_{\lambda+1}}\ldots A^{x_m}A^{x_0}\ldots A^{x_{\varkappa-1}}\,|\,A^{x_\varkappa}A^{x_\lambda})$$
$$= P(A^{x_\lambda}\ldots A^{x_m}A^{x_0}\ldots A^{x_\varkappa}) : P(A^{x_\varkappa}A^{x_\lambda})$$

ist, womit zugleich der erste Teil des Beweises geführt ist.

2) hinreichend: Es ist

$$P(A^{x_\nu}|A^{x_0}\dots A^{x_{\nu-1}}A^{x_{\nu+1}}\dots A^{x_m}) = \frac{P(A^{x_0}\dots A^{x_m})}{P(A^{x_0}\dots A^{x_{\nu-1}}A^{x_{\nu+1}}\dots A^{x_m})}.$$

Nach Voraussetzung ist der Zähler dieses Ausdrucks

$$= \frac{P(A^{x_{\nu-1}}A^{x_\nu}A^{x_{\nu+1}})\,P(A^{x_{\nu+1}}\dots A^{x_m}A^{x_0}\dots A^{x_{\nu-1}})}{P(A^{x_{\nu-1}}A^{x_{\nu+1}})}$$

und damit der ganze Ausdruck, wie in der Definition gefordert,

$$= P(A^{x_\nu}|A^{x_{\nu-1}}A^{x_{\nu+1}}).$$

Später wird gezeigt werden, daß ein NP durch Vorgabe der Wn

$$P(A^m_{\varkappa}|A^{m-1}_{\lambda}A^{m+1}_{i}) \qquad (m=0,1,\dots,n \text{ mit } A^{-1}_{i}=A^{n}_{i} \text{ und } A^{n+1}_{i}=A^{0}_{i}),$$

die noch gewissen Verträglichkeitsbedingungen genügen müssen, bestimmt ist. Aus diesen Größen lassen sich demnach die Wn beliebiger Ereignisse des NP berechnen. Dieses Ziel ist an sich erreicht, wenn die Wn $P(A^0_{i_0}A^1_{i_1}\dots A^n_{i_n})$ bekannt sind. Denn alle übrigen absoluten Wn setzen sich additiv aus diesen Größen zusammen.

Nun gilt allgemein

$$P(A^0A^1\dots A^n) = P(A^0A^n)\,P(A^1|A^0A^n)\,P(A^2|A^0A^1A^n)\dots P(A^{n-1}|A^0A^1\dots A^{n-2}A^n).$$

Für einen NP vereinfacht sich diese Formel nach Definition 1 zu

$$(1) \qquad P(A^0A^1\dots A^n) = P(A^0A^n)\prod_{m=1}^{n-1}P(A^m|A^{m-1}A^n).$$

§ 2. Offene Nachbarprozesse und Markoff-Ketten

In den Anwendungen interessieren vor allem solche NP, bei denen (nach eventuell passender zyklischer Umnumerierung der Stellen) das Ereignis A^0 von A^n nicht abhängt, wenn das Eintreten eines Ereignisses der Stellen $x=1,2,\dots,n-1$ bekannt ist. Solche NP seien offen genannt, und zwar deshalb, weil die Stellen 0 und n in diesem Fall eine Randstellung einnehmen, die durch Abtragen der Stellen $0,1,\dots,n$ auf einer offenen Kurve (Geradenstück) veranschaulicht werden kann.

Definition 2. *Ein NP heißt offen, wenn*

$$(1) \qquad P(A^0|A^xA^n) = P(A^0|A^x)$$

für jedes $x \neq 0, n$ *gilt*[4]).

Wie weitreichend diese Einschränkung ist, zeigt der folgende

Satz 1. *Für offene NP ist*

$$1)\quad P(A^{x_0}|A^{x_1}\dots A^{x_m}) = P(A^{x_0}|A^{x_1}),$$
$$2)\quad P(A^{x_m}|A^{x_0}\dots A^{x_{m-1}}) = P(A^{x_m}|A^{x_{m-1}})$$

[4]) In § 8 wird noch gezeigt, daß es genügt, (1) für irgendein $x \neq 0,n$ zu fordern.

jeweils für beliebige $x_0, x_1, \ldots, x_m$.

Beweis. Zu 1). Zunächst wird gezeigt, daß für offene NP

$$P(A^0 | A^{x_1} \ldots A^{x_m}) = P(A^0 | A^x) \qquad 0 < x_1 \tag{2}$$

gilt. Ist $x_m = n$, so folgt dies leicht nach § 1, Satz 1 und Definition 2 aus

$$P(A^0 | A^{x_1} \ldots A^n) = P(A^0 | A^{x_1} A^n) = P(A^0 | A^{x_1}).$$

Für $x_m \neq n$ gilt nach dem eben Bewiesenen

$$P(A^0 A_i^n | A^{x_1} \ldots A^{x_m}) = P(A_i^n | A^{x_1} \ldots A^{x_m})\, P(A^0 | A^{x_1}).$$

Und daraus folgt (2) durch Aufsummieren über i.

Wendet man (2) auf die Formel

$$\begin{aligned} P(A^0 A^{x_0} | A^{x_1} \ldots A^{x_m}) &= P(A^0 | A^{x_1} \ldots A^{x_m})\, P(A^{x_0} | A^0 A^{x_1} \ldots A^{x_m}) \\ &= P(A^{x_0} | A^{x_1} \ldots A^{x_m})\, P(A^0 | A^{x_0} \ldots A^{x_m}) \end{aligned}$$

an, so ergibt sich zusammen mit Definition 1

$$P(A^0 | A^{x_1})\, P(A^{x_0} | A^0 A^{x_1}) = P(A^{x_0} | A^{x_1} \ldots A^{x_m})\, P(A^0 | A^{x_0}).$$

Andererseits ist ebenfalls unter Beachtung von (2) $\left(P(A^0 A^{x_0} | A^{x_1}) = \right)$

$$P(A^0 | A^{x_1})\, P(A^{x_0} | A^0 A^{x_1}) = P(A^{x_0} | A^{x_1})\, P(A^0 | A^{x_0}).$$

Aus der Gleichheit der linken Seiten dieser beiden Gleichungen und

$$P(A^0 | A^{x_0}) \neq 0$$

folgt dann vollends die Gültigkeit von 1).

Zu 2). Nach § 1, Satz 1 ist zunächst

$$P(A^{x_m} | A^{x_0} \ldots A^{x_{m-1}}) = P(A^{x_m} | A^{x_0} A^{x_{m-1}}) = \frac{P(A^{x_0} A^{x_{m-1}} A^{x_m})}{P(A^{x_0} A^{x_{m-1}})}.$$

Unter Verwendung von 1) ergibt die weitere Rechnung

$$P(A^{x_m} | A^{x_0} \ldots A^{x_{m-1}}) = \frac{P(A^{x_0} | A^{x_{m-1}} A^{x_m})\, P(A^{x_{m-1}} A^{x_m})}{P(A^{x_0} | A^{x_{m-1}})\, P(A^{x_{m-1}})} = P(A^{x_m} | A^{x_{m-1}}),$$

wie gefordert.

Die 2. Aussage von Satz 1 ist identisch mit der Definition für alternative MK, die nach rechts fortschreiten. Also ist jeder offene NP eine MK. Andererseits ist nach Definition 1 jede MK ein NP, und da die Umkehrung einer MK wieder eine solche ist[5]), ist also für jede MK auch Definition 2 erfüllt. Also ist auch jede MK ein offener NP.

Trotz dieser Äquivalenz der beiden Begriffe sollen dieselben hier auseinandergehalten werden. Und zwar hinsichtlich der Art der vorzugebenden Größen, sowie des Unterschieds zwischen homogenen NP und homogenen MK und der bei offenen NP notwendig beschränkten Anzahl der Ereignisse.

[5]) Siehe J. DOOB [2], S. 83.

Wie schon in der Problemstellung erwähnt, ist bei den MK eine ausgezeichnete Richtung vorhanden. Dies folgt auch daraus, daß die Übergangs-Wn $P(A^m|A^{m+1})$ der Umkehrung einer MK im allgemeinen von den Anfangs-Wn der letzteren abhängig sind. Außerdem sind die Übergangs-Wn der Umkehrung einer homogenen MK — mit Ausnahme der stationären Kette — von der Stelle m abhängig. Bei den NP werden dagegen nur Wn vorgegeben, die keine Richtung bevorzugen. Nach der Definition der NP ist es naheliegend, die Größen $P(A^m|A^{m-1}A^{m+1})$ vorzugeben, die offenbar invariant gegenüber einer Richtungsänderung sind und außerdem nicht von den zusätzlich vorzugebenden Rand-Wn $P(A^0|A^1)$, $P(A^n|A^{n-1})$ abhängen.

Wie in § 1, sei auch für offene NP noch eine zweite Definition in Form einer Produktregel angegeben.

Satz 2. *Die Ereignisse $A^0, A^1, \ldots, A^n$ bilden genau dann einen offenen NP, wenn*

$$P(A^{x_0}\ldots A^{x_\nu}\ldots A^{x_m}) = \frac{P(A^{x_0}\ldots A^{x_\nu})\,P(A^{x_\nu}\ldots A^{x_m})}{P(A^{x_\nu})}$$

für beliebige $x_0, \ldots, x_\nu, \ldots, x_m$ mit $0 \leqq \nu \leqq m$ gilt.

Der Beweis ergibt sich leicht aus § 1, Satz 2, wenn dort $\varkappa = 0$ gesetzt wird.

Nach kurzer Zwischenrechnung ergibt sich hieraus die für die Berechnung offener NP wichtige Formel

$$(3)\quad \begin{cases} P(A^0A^1\ldots A^m\ldots A^n) = P(A^0|A^1)\,P(A^1|A^2)\ldots \\ \qquad\qquad \ldots P(A^{m-1}|A^m)\,P(A^m)\,P(A^{m+1}|A^m)\ldots P(A^n|A^{n-1}) \end{cases}$$

für $0 < m < n$ und sinngemäßer Abänderung für $m = 0, n$.

Abschließend sei noch vermerkt, daß die Ereignisse A^x im folgenden auch mit $A, B, C, \ldots$ und die dazu komplementären mit $\bar{A}, \bar{B}, \bar{C}, \ldots$ bezeichnet werden. Dabei soll jedoch in einer Beziehung, in der mehrere Ereignisse auftreten, dem Ereignis mit größerer oberer Indexzahl der in der alphabetischen Ordnung weiter rechts stehende Buchstabe zugeordnet sein.

§ 3. Das Wahrscheinlichkeitsverhältnis

Viele der abzuleitenden Gleichungen werden einfacher und übersichtlicher, wenn man neben der W $P(A)$, mit der ein Ereignis A eintritt, das Wahrscheinlichkeitsverhältnis (Wv)

$$Q(A) = \frac{P(A)}{P(\bar{A})}$$

benützt[6]). $P(A)$ und $Q(A)$ sind einander eineindeutig zugeordnet, falls man noch $Q(E) = \infty$ für $P(E) = 1$ setzt[7]). Denn es ist

$$(1)\qquad Q(A) = \frac{P(A)}{1 - P(A)} \quad \text{und} \quad P(A) = \frac{Q(A)}{1 + Q(A)}.$$

[6]) Ein weiterer Vorteil dieser Schreibweise wird im II. Kapitel ersichtlich.

[7]) Dabei gelten die üblichen Rechenregeln für ∞.

Ferner ist $Q(A) \geqq 0$, und für $P(A) < P(B)$ ist auch $Q(A) < Q(B)$. (Den Wn $P = 0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1$ entsprechen die Werte $Q = 0, \frac{1}{2}, 1, 2, \infty$.) Wichtig ist noch die Beziehung

$$Q(A) \cdot Q(\bar{A}) = 1,$$

die der Gleichung $P(A) + P(\bar{A}) = 1$ entspricht. Erwähnt sei noch, daß $Q(A)$ eine reelle positive, aber nicht additive Mengenfunktion des Ereignisraumes ist. Das Additionstheorem für $Q(A \cup B)$ ist dementsprechend komplizierter als das für Wn.

Ebenso sei noch das bedingte Wv $Q(B|A)$ definiert durch

$$Q(B|A) = \frac{P(B|A)}{P(\bar{B}|A)} = \frac{P(AB)}{P(A\bar{B})}. \tag{2}$$

Schließlich wird noch die Abkürzung

$$R(B, A) = \frac{Q(B|A)}{Q(B|\bar{A})}$$

verwendet.

§ 4. Zwei abhängige Ereignisse

In § 5 wird gezeigt werden, daß ein offener NP mit m Stellen auf einen offenen NP mit $m-1$ Stellen zurückgeführt werden kann. Dadurch gelangt man rekursiv auf den trivialen NP mit zwei Stellen. Um später nicht unterbrechen zu müssen, sei dieser Fall vorangestellt.

Die Problemstellung ist folgende: Die vier bedingten Wn

$$P(B|A),\ P(B|\bar{A}),\ P(A|B),\ P(A|\bar{B}) \tag{1}$$

zweier Ereignisse A und B seien gegeben. Gesucht sind die Wn $P(A)$, $P(B)$, $P(AB)$ usw.

Dabei wird im Einklang mit § 1 vorausgesetzt, daß alle bedingten Wn $\neq 0$ (und damit $\neq 1$) sind.

Nach (3.2) genügen $P(AB), \ldots$ dem homogenen Gleichungssystem

$$\begin{pmatrix} 1 & 0 & -Q(B|A) & 0 \\ 0 & 1 & 0 & -Q(B|\bar{A}) \\ 0 & 0 & 1 & -Q(A|\bar{B}) \\ 1 & -Q(A|B) & 0 & 0 \end{pmatrix} \begin{pmatrix} P(AB) \\ P(\bar{A}B) \\ P(A\bar{B}) \\ P(\bar{A}\bar{B}) \end{pmatrix} = 0.$$

Ist

$$\frac{Q(B|A)}{Q(B|\bar{A})} = \frac{Q(A|B)}{Q(A|\bar{B})}, \quad \text{also} \quad R(A, B) = R(B, A), \tag{2}$$

so hat das Gleichungssystem den Rang 3 und die Lösung

$$P(AB) : P(\bar{A}B) : P(A\bar{B}) : P(\bar{A}\bar{B}) = Q(A|\bar{B})\, Q(B|A) : Q(B|\bar{A}) : Q(A|\bar{B}) : 1.$$

Diese wird eindeutig durch Normieren auf Wn.

Außerdem ergibt sich

$$(3)\qquad \begin{cases} Q(A) = \dfrac{P(AB)+P(A\bar{B})}{P(\bar{A}B)+P(\bar{A}\bar{B})} = Q(A|\bar{B})\,\dfrac{1+Q(B|A)}{1+Q(B|\bar{A})} \\ \qquad = Q(A|B)\,\dfrac{1+Q(B|A)}{R(A,B)+Q(B|A)}. \end{cases}$$

Aus den vier Größen (1) können demnach die gefragten Wn eindeutig bestimmt werden, wenn sie der einzigen Verträglichkeitsbedingung (2) genügen.

§ 5. Rechenregeln und Verträglichkeitsbedingungen für offene NP

Ein offener NP, der an jeder der Stellen $x=0,1,2,\ldots,n$ die Werte 0 oder 1 annehmen kann, sei gegeben durch die Verkettungs-Wn $P(A^m|A^{m-1}A^{m+1})$ $(0<m<n)$ und die Rand-Wn $P(A^0|A^1)$ und $P(A^n|A^{n-1})$. Aus den Verkettungs-Wn und den rechten Rand-Wn lassen sich die Wn $P(A^{n-1}|A^{n-2})$ berechnen. Faßt man nun diese Größen als rechte Rand-Wn auf, so erhält man einen NP für die Stellen $x=0,1,\ldots,n-1$. Ebenso läßt sich der NP von links her abbauen. Reduziert man den NP in dieser Weise auf die Stellen $x=k$ und $k+1$ (oder k und $k-1$), so läßt sich hieraus nach (4.3) $P(A^k)$ berechnen.

Im folgenden werden nun die hierzu nötigen Rekursionsformeln und Verträglichkeitsbedingungen hergeleitet.

Nach (3.2) gilt für drei *beliebige* Ereignisse A, B, C:

$$(1)\qquad Q(B|AC) = \frac{P(ABC)}{P(A\bar{B}C)} = \frac{P(C|AB)}{P(C|A\bar{B})}\cdot Q(B|A),$$

$$(2)\qquad \frac{Q(A|BC)}{Q(A|\bar{B}C)} = \frac{Q(B|AC)}{Q(B|\bar{A}C)} \quad \text{und} \quad \frac{Q(B|AC)}{Q(B|A\bar{C})} = \frac{Q(C|AB)}{Q(C|A\bar{B})}.$$

Ersetzt man die Wn in (1) durch die Wv, so erhält man

$$(3)\qquad \begin{cases} Q(B|A) = Q(B|AC)\,\dfrac{Q(C|A\bar{B})}{Q(C|AB)}\,\dfrac{1+Q(C|AB)}{1+Q(C|A\bar{B})} \\ \qquad = \dfrac{1+Q(C|AB)}{Q(\bar{B}|A\bar{C})+Q(\bar{B}|AC)\,Q(C|AB)}. \end{cases}$$

Sind A, B, C Ereignisse eines offenen NP, die den Stellen $x_1<x_2<x_3$ zugeordnet sind, so vereinfachen sich (1), (2) und (3) wegen $Q(C|AB)=Q(C|B)$ zu

$$(4)\qquad Q(B|A) = Q(B|AC)\,\frac{P(C|\bar{B})}{P(C|B)},$$

$$(5)\qquad R(A,B) = \frac{Q(B|AC)}{Q(B|\bar{A}C)}, \qquad R(B,C) = \frac{Q(B|AC)}{Q(B|A\bar{C})},$$

$$(6)\qquad Q(B|A) = \frac{1+Q(C|B)}{Q(\bar{B}|A\bar{C})+Q(\bar{B}|AC)\,Q(C|B)}.$$

Ebenso erhält man durch Vertauschen von A und C

$$(7)\qquad Q(B|C) = \frac{1+Q(A|B)}{Q(\bar{B}|\bar{A}C)+Q(\bar{B}|AC)\,Q(A|B)}$$

und aus jeder der Beziehungen (5)

$$\begin{vmatrix} Q(B|AC) & Q(B|\bar{A}C) \\ Q(B|A\bar{C}) & Q(B|\bar{A}\bar{C}) \end{vmatrix} = 0. \tag{8}$$

Faßt man A, B, C als drei aufeinanderfolgende Ereignisse A^{m-1}, A^m, A^{m+1} auf, so sind (6) und (7) die gesuchten Rekursionsformeln für die Übergangs-Wv. Führt man noch die Abkürzungen

$$\begin{cases} \alpha_m = \dfrac{1}{a_m} = Q(A_1^m | A_1^{m-1} A_1^{m+1}) \\ \beta_m = \dfrac{1}{b_m} = Q(A_1^m | A_0^{m-1} A_1^{m+1}) \\ \gamma_m = \dfrac{1}{c_m} = Q(A_1^m | A_1^{m-1} A_0^{m+1}) \\ \delta_m = Q(A_1^m | A_0^{m-1} A_0^{m+1}), \end{cases} \tag{9}$$

sowie

$$l_m = Q(A_1^m | A_1^{m+1}), \quad r_m = Q(A_1^m | A_1^{m-1})$$
$$l'_m = Q(A_1^m | A_0^{m+1}), \quad r'_m = Q(A_1^m | A_0^{m-1})$$

ein, so folgt aus (6) und (7)

$$\begin{cases} r_m = \dfrac{1 + r_{m+1}}{c_m + a_m r_{m+1}}, \quad l_m = \dfrac{1 + l_{m-1}}{b_m + a_m l_{m-1}} \\ \text{oder} \\ r_m = L_m(r_{m+1}), \qquad l_m = M_m(l_{m-1}). \end{cases} \tag{10}$$

r_m bzw. l_m ist also eine linear gebrochene Funktion von r_{m+1} bzw. l_{m-1}. Berechnet man nun aus den gegebenen Wn nach (3.1) die Wv

$$\alpha_m, \beta_m, \gamma_m, \delta_m \ (m = 1, 2, \ldots, n-1), \ l_0, l'_0, r_n, r'_n \tag{11}$$

oder nimmt am besten diese Wv als gegeben an, so lassen sich r_m und l_m rekursiv aus (10) berechnen. r'_m und l'_m erhält man aus den aus (5) folgenden Beziehungen

$$\frac{r_m}{r'_m} = \frac{\alpha_{m-1}}{\gamma_{m-1}} = \frac{\alpha_m}{\beta_m} \quad \text{und} \quad \frac{l_m}{l'_m} = \frac{\alpha_m}{\gamma_m} = \frac{\alpha_{m+1}}{\beta_{m+1}}. \tag{12}$$

Daraus erhält man nach (4.3) vollends $Q(A_1^m)$. Insbesondere folgt hieraus für $0 < m < n$

$$Q(A_1^m) = l_m \frac{1 + r_{m+1}}{\dfrac{\alpha_m}{\gamma_m} + r_{m+1}} = l_m a_m \frac{1 + r_{m+1}}{c_m + a_m r_{m+1}}$$

und daraus nach (10)

$$Q(A_1^m) = l_m a_m r_m = \frac{Q(A_1^m | A_1^{m-1}) \, Q(A_1^m | A_1^{m+1})}{Q(A_1^m | A_1^{m-1} A_1^{m+1})}. \tag{13}$$

Auf die Berechnung von Wn anderer Ereignisse sei hier nicht eingegangen, zumal hierfür durch (2.3) ein Weg angegeben ist.

Die Verträglichkeitsbedingungen, die die gegebenen Wv (11) eines offenen NP erfüllen müssen, folgen leicht aus (8) und (12) und seien in dem folgenden Satz zusammengestellt:

Die Größen (11) *können genau dann als Verkettungs- und Rand-Wv eines offenen NP mit* $n+1$ *Stellen aufgefaßt werden und bestimmen diesen eindeutig, wenn sie endlich und* >0 *sind und den folgenden vier Verträglichkeitsbedingungen genügen:*

$$\text{I.}\quad \begin{vmatrix}\alpha_m & \beta_m\\ \gamma_m & \delta_m\end{vmatrix}=0 \quad (0<m<n), \qquad \text{III.}\quad \begin{vmatrix}l_0 & \alpha_1\\ l_0' & \beta_1\end{vmatrix}=0,$$

$$\text{II.}\quad \begin{vmatrix}\alpha_m & \alpha_{m+1}\\ \gamma_m & \beta_{m+1}\end{vmatrix}=0 \quad (0<m<n-1), \qquad \text{IV.}\quad \begin{vmatrix}\alpha_{n-1} & r_n\\ \gamma_{n-1} & r_n'\end{vmatrix}=0.$$

Beweis. 1) Daß diese Bedingungen notwendig sind, folgt für I aus (8) und für II, III, IV aus (12).

2) Umgekehrt überzeugt man sich leicht, daß etwa die Größen

$$\alpha_1,\alpha_2,\ldots,\alpha_{n-1};\ \beta_1,\beta_2,\ldots,\beta_{n-1};\ l_0,l_0';\ r_n \tag{14}$$

beliebig vorgegeben werden können, ohne die Verträglichkeitsbedingungen zu verletzen. Weiter können daraus mit Hilfe von (10) $r_{n-1}, r_{n-2}, \ldots, r_1$ und mit Hilfe von (12) $r_n', r_{n-1}', \ldots, r_2'$ berechnet werden. Schließlich ergibt sich aus l_0, l_0', r_1 nach § 4 r_1' und $Q(A_1^0)$. Andererseits ist aber durch willkürliche Vorgabe der Größen $Q(A_1^0), r_1, \ldots, r_n, r_1', \ldots, r_n'$ eine MK bestimmt. Die Größen (14) bestimmen also stets einen offenen NP und die restlichen Größen von (11) ergeben sich eindeutig aus den Verträglichkeitsbedingungen. Gäbe es einen zweiten, davon verschiedenen NP mit den Wv (14), so wäre notwendig mindestens eine der Größen $Q(A_1^0), r_m, r_m'$ von den eben erwähnten verschieden und die Eineindeutigkeit der benützten Rechenregeln würde zwangsweise zu einem Widerspruch führen.

Aus dem Konstruktionsprinzip des 2. Teils dieses Beweises kann nun leicht die Zahl der unabhängig vorgebbaren Wn eines NP angegeben werden. Nach (14) sind dies bei einem NP mit $n+1$ Stellen $2n+1$ Größen, etwa $2n-2$ Verkettungs-Wn und drei Rand-Wn. Das sind natürlich genau so viel wie bei einer nichthomogenen MK, bei der eine Anfangs-W und $2n$ Übergangs-Wn gegeben sind.

Dagegen sind bei den homogenen NP, die im nächsten Paragraphen behandelt werden, vier Größen frei wählbar, während homogene MK schon durch drei bestimmt sind.

§ 6. Homogene offene Nachbarprozesse

Ein NP heiße homogen, wenn seine Verkettungs-Wn $P(A^m|A^{m-1}A^{m+1})$ von der Stelle m unabhängig sind. Aus (5.4) folgt sofort, daß jede homogene MK ein homogener offener NP ist. Das Umgekehrte ist nicht der Fall, wie man sich leicht an einem Beispiel klar macht. Jedoch wird im nächsten Paragraphen bewiesen, daß bei beliebiger Vorgabe der Verkettungs-Wn eines homogenen offenen NP die Rand-Wn stets so gewählt werden können, daß der NP eine homogene MK bildet.

Für das Folgende sollen also nun $\alpha_m, \beta_m, \gamma_m, \delta_m$ unabhängig von m sein, und es werde

$$\alpha_m = \alpha = \frac{1}{a}; \quad \beta_m = \beta = \frac{1}{b}; \quad \gamma_m = \gamma, \quad \delta_m = \delta$$

gesetzt.

Die Verträglichkeitsbedingungen für die Verkettungs- und Rand-Wv $\alpha, \beta, \gamma, \delta; l_0, l_0', r_n, r_n'$ homogener offener NP ergeben sich als Sonderfall von (5. I—IV) und lauten

$$\text{I}'. \begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} = 0, \qquad \text{II}'. \ \beta = \gamma, \qquad \text{III}'. \begin{vmatrix} l_0 & \alpha \\ l_0' & \beta \end{vmatrix} = 0, \qquad \text{IV}'. \begin{vmatrix} \alpha & r_n \\ \beta & r_n' \end{vmatrix} = 0.$$

Vorgebbar sind im homogenen Fall vier Größen. Für den jetzigen Paragraphen sei angenommen, daß α, β, l_0 und r_n gegeben seien. Diese vier Wv können, wie man sich leicht überzeugt, beliebig (natürlich >0) gewählt werden, ohne die Verträglichkeitsbedingungen zu verletzen.

Die Rekursionsformeln für die Übergangs-Wv lauten in diesem Fall nach (5.10)

$$r_{m-1} = \frac{1 + r_m}{b + a r_m} \quad \text{und} \quad l_{m+1} = \frac{1 + l_m}{b + a l_m}$$

oder, falls man noch $L(x) = \frac{1 + x}{b + a x}$ einführt,

$$r_{m-1} = L(r_m) \quad \text{und} \quad l_{m+1} = L(l_m).$$

Setzt man $L\big(L^{n-1}(x)\big) = L^n(x)$ und $L^0(x) = x$, so ist

$$(1) \qquad r_m = L^{n-m}(r_n) \quad \text{und} \quad l_m = L^m(l_0).$$

Dabei ist $L^p(x)$ wieder linear gebrochen in x.

Hieraus erkennt man überdies, daß $Q(A^m | A^{m-1})$ bzw. $Q(A^m | A^{m+1})$ von dem linken bzw. rechten Rand-Wv nicht abhängt.

Ferner ist nach (5.12)

$$(2) \qquad \frac{r_p}{r_p'} = \frac{l_m}{l_m'} = \frac{\alpha}{\beta} = R(A_1^m, A_1^{m-1}) = \text{const} \qquad \text{für } m = 1, 2, \ldots, n.$$

Hieraus erhält man nach (4.3)

$$Q(A_1^0) = l_0 \frac{1 + L^{n-1}(r_n)}{\frac{\alpha}{\beta} + L^{n-1}(r_n)} = l_0 a L^n(r_n) \quad \text{und} \quad Q(A_1^n) = L^n(l_0) a r_n.$$

Wegen (5.13) gilt also

$$Q(A_1^m) = L^m(l_0) a L^{n-m}(r_n), \quad \text{für } m = 0, 1, \ldots, n.$$

Natürlich kann $Q(A_1^m)$ nach Berechnung der r_p, r_p' und $Q(A_1^0)$ rekursiv auch nach den Rechenregeln für nichthomogene MK ermittelt werden. Der angegebene Weg ist jedoch einfacher, zumal anschließend eine explizite Darstellung von $L^m(x)$ angegeben wird.

Bei vielen Fragen, wie etwa bei der Berechnung von Mittelwerten, Streuungen, Verteilungsfunktionen bei seltenen Ereignissen usw., ist es zweckmäßiger, von den Wn $P(A^m | A^0 A^n)$, die von den Rand-Wn unabhängig sind, auszugehen. Diese Wn lassen sich im homogenen Fall, wie in § 8 gezeigt werden wird, ebenfalls in geschlossener Form angeben.

§ 7. Einige Eigenschaften der homogenen offenen NP

Eine explizite Darstellung von $L^m(x)$ ergibt sich leicht mit Hilfe des Fixpunktsatzes linearer Funktionen. Danach läßt sich $y = L(x) = \frac{1+x}{b+ax}$ mit $a \neq b$ auf die Form

$$\frac{y-\xi_1}{y-\xi_2} = \varrho \frac{x-\xi_1}{x-\xi_2} \tag{1}$$

bringen, wenn die Lösungen ξ_1, ξ_2 der quadratischen Gleichung $x = L(x)$ voneinander verschieden sind. Dies ist stets der Fall; denn wegen $a, b > 0$ ist $\xi_1 \cdot \xi_2 < 0$. Ferner ist $\varrho = (a\xi_1 - 1) : (a\xi_2 - 1)$ mit $|\varrho| < 1$, falls man unter ξ_1 die positive der Wurzeln versteht.

Unter Beachtung von $y = L^p(x) = L\left(L^{p-1}(x)\right)$ erhält man nun aus (1) leicht die Darstellung

$$\frac{L^m(x)-\xi_1}{L^m(x)-\xi_2} = \varrho^m \frac{x-\xi_1}{x-\xi_2}. \tag{2}$$

Es erübrigt sich also, die Übergangs-Wv $r_m = L^{n-m}(r_n)$ und $l_m = L^m(l_0)$ rekursiv zu berechnen.

Außerdem folgt aus (2) wegen $|\varrho| < 1$

$$\lim_{m \to \infty} L^m(x) = \xi_1 \quad \text{für } x \geqq 0.$$

Für die Übergangs-Wv folgt hieraus und aus (6.2):

Es existiert $\lim_{m \to \infty} Q(A_1^m | A_i^{m+1})$, und der Grenzwert ist unabhängig von den Rand-Wn. Aus (6.1) folgt, daß diese Aussagen auch für $Q(A_\varkappa^k | A_i^{k-1})$ bei festem k und $n \to \infty$ zutreffen.

Demnach verhält sich ein homogener offener NP für genügend großes n in der Nähe des rechten (linken) Randes wie eine homogene MK, die nach links (rechts) fortschreitet. Diese Ketten haben gleiche Übergangs-Wn, jedoch im allgemeinen verschiedene Anfangs-Wn.

Zu erwähnen ist noch der Fall $a = b$ ($\alpha = \beta$), der sich jedoch dem allgemeinen Fall trivialerweise unterordnet. Hier ist $L(x) \equiv 1/a = \alpha$ und wegen (6.I', II') gilt $Q(A_1^m | A_\iota^{m-1} A_\lambda^{m+1}) = \alpha$ (unabhängig von ι, λ). Demnach ist $A_\varkappa^m$ von A_ι^{m-1} und A_λ^{m+1} und damit von allen übrigen Ereignissen A_ϱ^r unabhängig, und es ist $Q(A_1^m) = \alpha$ für $m \neq 0, n$. Der NP ist also für $\alpha = \beta$ nicht verkettet und kann deshalb von den weiteren Betrachtungen ausgeschlossen bleiben.

$Q(A^0)$ und $Q(A^n)$ als gegebene Rand-Wn. Mit Hilfe der expliziten Darstellung von $L^m(x)$ läßt sich die Berechnung des NP, bei dem die Größen $Q(A_1^0)$, $Q(A_1^n)$, α und β gegeben sind, auf die des bisher behandelten Falles zurückführen. Dies wird erreicht, indem man aus diesen vier Größen l_0 und r_n berechnet.

Hierzu benützt man die in § 6 abgeleiteten Beziehungen

$$(3) \qquad Q(A_1^0) = a l_0 L^n(r_n) \quad \text{und} \quad Q(A_1^n) = a r_n L^n(l_0).$$

Aus diesen beiden Gleichungen sind nun l_0 und r_n zu ermitteln, etwa durch Auflösung der Gleichung

$$Q(A_1^0) = a l_0 L^n\left(\frac{Q(A_1^n)}{a L^n(l_0)}\right).$$

Bringt man diese wegen der Gruppeneigenschaft linear gebrochener Funktionen in l_0 quadratische Gleichung mit Hilfe von (2) auf die übliche Form, so erkennt man bei Beachtung des Vietaschen Satzes, daß genau eine positive Lösung l_0 und damit auch genau ein zugehöriges positives r_n existiert.

Eine zweite Möglichkeit der Berechnung eines NP mit vorgegebenen absoluten Rand-Wn, die auch im inhomogenen Fall angewendet werden kann, sei ebenfalls kurz skizziert. Aus (5.1) folgt für $A = A^0$, $B = A^n$, $C = A^m$, daß sich $R(A^n, A^0)$ durch die Wn $P(A^m | A^0 A^n)$ ausdrücken läßt. Ein Weg zur Berechnung dieser Wn, die nur von α und β abhängen, wird in § 8 beschrieben.

Aus $R(A^n, A^0)$, $Q(A^0)$, $Q(A^n)$ erhält man aber nach § 3 die Größen $P(A^0 A^n)$ und damit schließlich etwa $P(A_\varkappa^m) = \sum_\iota \sum_\lambda P(A_\varkappa^m | A_\iota^0 A_\lambda^n)\, P(A_\iota^0 A_\lambda^n)$.

Erwähnt sei noch, daß bei Vorgabe einer absoluten und einer bedingten Rand-W $l_{n-1}, l_{n-2}, \ldots, l_1$ bzw. $r_1, r_2, \ldots, r_{n-1}$ nach (6.1) berechnet werden können, womit der NP auf eine nichthomogene MK zurückgeführt ist. Einen zweiten Weg zur Berechnung des NP bieten die Formeln (3) an.

Wann ist ein offener NP eine homogene MK? Diese schon in § 6 angeschnittene Frage läßt sich jetzt sehr leicht beantworten. Da wegen (5.4) jede homogene MK ein homogener offener NP ist, kann umgekehrt nur ein solcher eine homogene MK sein. Ein homogener offener NP ist aber genau dann eine homogene MK nach rechts, wenn $r_1 = r_2 = \cdots = r_n = r$ ist. Denn dies bedeutet, daß $P(A_1^m | A_\iota^{m-1})$ konstant bezüglich m ist. Aus (6.1) erhält man dafür die Bedingung $r = L(r)$ mit $r \geqq 0$ oder äquivalent hierzu

$$(4) \qquad r_n = \frac{1 + r_n}{b + a r_n}.$$

Besteht also zwischen den Verkettungs-Wv a und b und des rechten Rand-Wv r_n eines NP die Relation (4), so ist derselbe eine homogene MK nach rechts. Hält man a und b fest, so ist $r_n = \xi_1$ die einzige positive Lösung von (4). Das Entsprechende gilt für eine nach links fortschreitende MK. Man erhält somit den Satz:

Der durch α, β, l_0, r_n bestimmte homogene offene NP ist genau dann eine homogene MK nach rechts (links), wenn $r_n = \xi_1$ ($l_0 = \xi_1$) ist. Ist beides der Fall, so ist der NP eine stationäre MK, d.h. es ist auch $P(A_\iota^m) = const.$

§ 8. Geschlossene Nachbarprozesse

Geschlossen soll ein NP heißen, wenn Definition 2 weder für die dort ausgezeichnete Stelle $x = 0$ noch für ein anderes x bei sinngemäßer zyklischer Vertauschung der x_m erfüllt ist.

Die Stellen $x = 0, 1, \ldots, n$ denkt man sich bei den geschlossenen NP am besten gleichmäßig auf einem Kreise verteilt. Ferner sei A^x für alle ganzzahligen x erklärt durch $A^x = A^y$ mit $x = y + k(n+1)$ (k ganz). Jedoch unterliegen die Stellen x der Ereignisse A^x, die in den folgenden Beziehungen auftreten, alle der Einschränkung $m \leqq x \leqq m + n + 1$ (m ganz).

Die Problemstellung ist folgende: Gegeben seien die Wv $Q(A_1^m | A_i^{m-1} A_x^{m+1})$, oder in der Schreibweise (5.9)

$$\alpha_m, \beta_m, \gamma_m, \delta_m > 0 \qquad m = 0, 1, \ldots, n, \tag{1}$$

eines NP. Gesucht sind die Wn $P(A_\mu^m)$, $P(A_\varkappa^k A_\lambda^l)$ usw. Dabei ist zu untersuchen, welche der Größen (1) unabhängig vorgegeben werden können, und ob sie andererseits zur vollständigen Bestimmung des NP ausreichen. Vorweg ist noch zu sagen, daß die folgenden Ableitungen auch für offene NP gelten. Jedoch ist es bei diesen bequemer, den oben eingeschlagenen Weg zu wählen. Ferner soll dieser Paragraph in erster Linie als Vorbereitung für Kapitel II aufgefaßt werden. Es wird deshalb meist auf eine Deutung und Spezialisierung der Ergebnisse etwa im Hinblick auf asymptotisches Verhalten oder auf homogene NP verzichtet.

Das Wesentliche der folgenden Ableitungen ist, aus den Größen (1) die Wv $Q(A^{x_1} | A^{x_2} A^{x_3})$ zu berechnen. Dies wird wieder in rekursiver Weise geschehen.

Zunächst gilt für vier Ereignisse A, B, C, D die den Stellen $x_1 < x_2 < x_3 < x_4$ zugeordnet seien:

$$P(BC|AD) = P(B|AD)\, P(C|ABD) = P(C|AD)\, P(B|ACD).$$

Nach § 1, Satz 1 folgt hieraus für NP

$$P(B|AD)\, P(C|BD) = P(C|AD)\, P(B|AC). \tag{2}$$

Dividiert man (2) durch die aus (2) durch Ersetzen von B durch $\bar{B}$ bzw. C durch $\bar{C}$ entstehende Gleichung, so erhält man

$$Q(B|AD) = Q(B|AC)\,\frac{P(C|\bar{B}D)}{P(C|BD)}, \qquad Q(C|AD) = \frac{P(B|A\bar{C})}{P(B|AC)}\, Q(C|BD). \tag{3}$$

Auf dieselbe Weise erhält man aus (3)

$$\frac{Q(B|AC)}{Q(B|A\bar{C})} = \frac{Q(C|BD)}{Q(C|\bar{B}D)}, \qquad \frac{Q(B|AC)}{Q(B|A\bar{C})} = \frac{Q(B|\bar{A}C)}{Q(B|\bar{A}\bar{C})}, \tag{4}$$

$$\frac{Q(B|AC)}{Q(\bar{B}|AC)} = \frac{Q(B|AD)}{Q(B|\bar{A}D)}, \qquad \frac{Q(C|BD)}{Q(C|B\bar{D})} = \frac{Q(C|AD)}{Q(C|A\bar{D})}. \tag{5}$$

Es sei noch darauf hingewiesen, daß aus den Gln. (3) umgekehrt die Gl. (2) folgt, was sich leicht durch Einsetzen von (3) in (2) — am besten mit Hilfe von (4) — zeigen läßt.

Ersetzt man in (3) die Wn durch ihre Wv, so erhält man unter Verwendung von (4) die Gleichungen

$$\begin{cases} Q(B|AD) = \dfrac{1 + Q(C|BD)}{Q(\bar{B}|A\bar{C}) + Q(\bar{B}|AC)\, Q(C|BD)} \\ Q(C|AD) = \dfrac{1 + Q(B|AC)}{Q(\bar{C}|\bar{B}D) + Q(\bar{C}|BD)\, Q(B|AC)}. \end{cases} \tag{6}$$

Damit sind die benötigten Gleichungen zusammengestellt. Alle weiteren Ergebnisse folgen durch Spezialisierung derselben.

Zunächst erhält man aus (4) für Ereignisse unmittelbar aufeinanderfolgender Stellen die Verträglichkeitsbedingungen

$$(7) \qquad \begin{vmatrix} \alpha_m & \alpha_{m+1} \\ \gamma_m & \beta_{m+1} \end{vmatrix} = 0 \qquad \begin{vmatrix} \alpha_m & \beta_m \\ \gamma_m & \delta_m \end{vmatrix} = 0 \qquad m = 0, 1, 2, \ldots, n.$$

Dabei sei, wie oben vereinbart, $\alpha_{n+1} = \alpha_0$, $\beta_{n+1} = \beta_0$.

Durch Vorgabe von $\alpha_0, \ldots, \alpha_n, \beta_0, \ldots, \beta_n$ sind also die Größen γ_m und δ_m eindeutig bestimmt.

Dann folgen aus (6) für die Ereignisse A_1^{m-1}, A_1^m, A_1^{m+1}, A_ϱ^r bzw. A_λ^l, A_1^{m-1}, A_1^m, A_1^{m+1} mit $r > m+1$, $l < m-1$ die gesuchten Rekursionsformeln

$$(8) \qquad Q(A_1^m | A_1^{m-1} A_\varrho^r) = \frac{1 + Q(A_1^{m+1} | A_1^m A_\varrho^r)}{c_m + a_m Q(A_1^{m+1} | A_1^m A_\varrho^r)} = L_m\big(Q(A_1^{m+1} | A_1^m A_\varrho^r)\big),$$

$$(9) \qquad Q(A_1^m | A_\lambda^l A_1^{m+1}) = \frac{1 + Q(A_1^{m-1} | A_\lambda^l A_1^m)}{b_m + a_m Q(A_1^{m-1} | A_\lambda^l A_1^m)} = M_m\big(Q(A_1^{m-1} | A_\lambda^l A_1^m)\big).$$

Dabei sind L_m und M_m die in § 5 eingeführten linear gebrochenen Funktionen.

Aus (5) folgt noch

$$(10) \qquad \begin{cases} Q(A_1^m | A_0^{m-1} A_\varrho^r) = \dfrac{a_m}{b_m} Q(A_1^m | A_1^{m-1} A_\varrho^r) \\ Q(A_1^m | A_\lambda^l A_0^{m+1}) = \dfrac{a_m}{c_m} Q(A_1^m | A_\lambda^l A_1^{m+1}). \end{cases}$$

Mit Hilfe dieser Rekursionsformeln läßt sich ein NP von $n+1$ Stellen auf einen NP von n Stellen reduzieren. Dieser Gedankengang soll aber hier nicht weiter verfolgt werden. Erwähnt sei jedoch noch, daß sich hierdurch relativ einfach ein Beweis für die Vollständigkeit der Verträglichkeitsbedingungen (7) – mit anderen Worten: für die Unabhängigkeit der Wv α_m, β_m – erbringen läßt.

Zur Berechnung der absoluten Wn des NP soll hier noch die Formel (5.3) benützt werden. Für $A = A_1^n$, $B = A_1^0$, $C = A_1^1$ ergibt sich hieraus

$$Q(A_1^0 | A_1^n) = L_0\big(Q(A_1^1 | A_1^0 A_1^n)\big),$$

was formal auch aus (8) für $m = 0$, $r = n$ und $\varrho = 1$ folgt. Ebenso erhält man

$$Q(A_1^0 | A_0^n) = \frac{a_0}{b_0} L_0\big(Q(A_1^1 | A_1^0 A_0^n)\big)$$

und

$$Q(A_1^n | A_1^0) = M_n\big(Q(A_1^{n-1} | A_1^0 A_1^n)\big).$$

Beachtet man schließlich noch, daß nach (8) und (9)

$$(11) \qquad Q(A_1^m | A_1^{m-1} A_\nu^n) = L_m L_{m+1} \ldots L_{n-2} Q(A_1^{n-1} | A_1^{n-2} A_\nu^n)$$

und

$$Q(A_1^m | A_\varkappa^0 A_1^{m+1}) = M_m M_{m-1} \ldots M_2 Q(A_1^1 | A_\varkappa^0 A_1^2)$$

ist, wobei für die linear gebrochenen Funktionen $F_p(F_q(x)) = F_p F_q x$ gesetzt wurde, so ergibt sich

$$Q(A_1^0|A_1^n) = L_0 L_1 \dots L_{n-2}\alpha_{n-1}$$
$$Q(A_1^0|A_0^n) = \frac{a_0}{b_0} L_0 L_1 \dots L_{n-2}\beta_{n-1}$$
$$Q(A_1^n|A_1^0) = M_n M_{n-1} \dots M_2 \alpha_1 .$$

Aus diesen drei Größen lassen sich nach § 4 $P(A_i^0)$, $P(A_\varkappa^n)$, $P(A_i^0 A_\varkappa^n)$ berechnen.

Aus $P(A_i^0 A_\varkappa^n)$, sowie den nach (11) und (10) berechenbaren Wn $P(A_1^m|A_i^{m-1}A_\varkappa^n)$ lassen sich aber nach (1.1) alle übrigen absoluten Wn ableiten.

Zur praktischen Berechnung der Wn $P(A_1^m)$, $P(A_\mu^m A_0^\nu)$ usw. ist (1.1) natürlich nicht geeignet. Hierfür gibt es bequemere Wege. So kann man $P(A_1^m)$ leicht aus $P(A_1^m|A_i^0 A_\varkappa^n)$ und $P(A_i^0 A_\varkappa^n)$ berechnen. Denn es ist $P(A_1^m) = \sum_i \sum_\varkappa P(A_1^m|A_i^0 A_\varkappa^n)\, P(A_i^0 A_\varkappa^n)$. $P(A_1^m|A_i^0 A_\varkappa^n)$ schließlich ist leicht aus den Größen $Q(A_i^{m+1}|A_\mu^m A_\nu^n)$ und $Q(A_\mu^m|A_\varkappa^0 A_\lambda^{m+1})$, die man zur Berechnung von $P(A_i^0 A_\varkappa^n)$ ohnedies braucht, zu ermitteln; und zwar ist nach (6)

$$Q(A_1^m|A_i^0 A_\varkappa^n) = \frac{1 + Q(A_1^{m+1}|A_1^m A_\varkappa^n)}{Q(A_0^m|A_i^0 A_0^{m+1}) + Q(A_0^m|A_i^0 A_1^{m+1})\, Q(A_1^{m+1}|A_1^m A_\varkappa^n)} .$$

Dieser Ausdruck läßt sich bei Benützung von (8) und (10) noch auf die einfache Gestalt

$$Q(A_1^m|A_i^0 A_\varkappa^n) = Q(A_1^m|A_i^0 A_1^{m+1}) \cdot a_m \cdot Q(A_1^m|A_1^{m-1}A_\varkappa^n) \tag{12}$$

bringen.

Nun noch ein Wort zu den offenen NP. Bei diesen ist

$$\alpha_0 = Q(A_1^0|A_1^n A_1^1) = Q(A_1^0|A_1^1) = l_0$$
$$\beta_0 = Q(A_1^0|A_0^n A_1^1) = Q(A_1^0|A_1^1) = l_0 .$$

Es muß also $\alpha_0 = \beta_0$ sein. Ebenso ist

$$\alpha_n = Q(A_1^n|A_1^{n-1}A_1^0) = Q(A_1^n|A_1^{n-1}) = r_n$$
$$\gamma_n = Q(A_1^n|A_1^{n-1}A_0^0) = Q(A_1^n|A_1^{n-1}) = r_n .$$

Also ist auch $\alpha_n = \gamma_n$. Dies folgt aber schon aus $\alpha_0 = \beta_0$ und der Verträglichkeitsbedingung (7) für $m = n$.

Umgekehrt folgt aus der Beziehung

$$\frac{Q(A_i^0|A_1^n A_\lambda^p)}{Q(A_i^0|A_0^n A_\lambda^p)} = \frac{Q(A_i^0|A_1^n A_\varkappa^p)}{Q(A_i^0|A_0^n A_\varkappa^p)} \qquad 0 < p < n, \quad 0 < q < n,$$

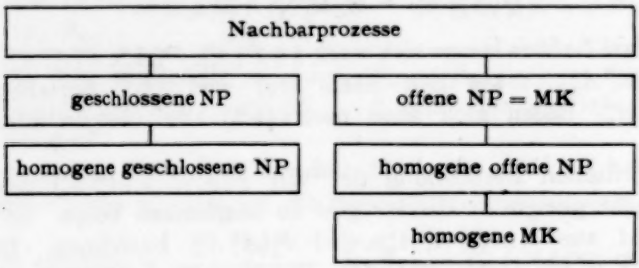
die sich aus (5) ergibt, daß aus

$$Q(A_i^0|A_\lambda^q A_1^n) = Q(A_i^0|A_\lambda^q A_0^n) = Q(A_i^0|A_\lambda^q) \qquad \text{für irgendein } q$$
$$Q(A_i^0|A_\varkappa^p A_\nu^n) = Q(A_i^0|A_\varkappa^p) \qquad \text{für alle } p = 1, 2, \dots, n-1$$

folgt. Dies war zur Definition 2 von § 2 noch nachzutragen.

Außerdem sieht man hieraus, daß der NP ein offener ist, wenn $\alpha_0 = \beta_0$ oder allgemein $\alpha_\nu = \beta_\nu$ ist. Ist $\alpha_\nu = \beta_\nu$ für zwei oder mehr Werte ν, so zerfällt der NP in zwei oder mehr offene NP.

Schließlich sei hier noch die in der Problemstellung erwähnte Übersicht über die verschiedenen Klassen von NP nachgetragen.



Hierbei ist jede Klasse von Prozessen in der darüberstehenden echt enthalten. Nicht übereinanderstehende Klassen sind fremd.

II. KAPITEL

Nachbarprozesse mit stetiger Ortskoordinate

§ 1. Vorbemerkungen und Definition der stetigen NP

Während bei den diskreten NP endlich vielen linear angeordneten Punkten je ein alternatives Ereignis zugeordnet wurde, soll bei den im folgenden betrachteten NP mit stetiger Ortskoordinate (kurz stetiger NP genannt) jedem Punkt x einer geschlossenen oder offenen Kurve ($0 \leqq x \leqq d$) ein Alternativereignis A_ι^x ($\iota = 0, 1$) zugeordnet werden. Der letztere Fall soll vorerst ausgeschlossen bleiben. Später — bei Berücksichtigung der Randverhältnisse — wird gezeigt, daß er sich dem Fall der geschlossenen Kurven unterordnet. Ferner sei A^x wieder für alle x erklärt durch

$$A^x = A^{x+nd} \qquad (n \text{ ganz}). \tag{1}$$

Eine Gesamtbeobachtung ist demnach eine d-periodische Funktion $f(x)$, die nur die Werte 0 und 1 annimmt. Der Ereignisraum bestehe etwa aus der Gesamtheit dieser Funktionen $f(x)$.

Ein hierzugehöriger stochastischer Prozeß soll stetiger NP heißen, wenn

$$P(A^{x_m} \mid A^{x_1} \dots A^{x_{m-1}} A^{x_{m+1}} \dots A^{x_n}) = P(A^{x_m} \mid A^{x_{m-1}} A^{x_{m+1}}) \tag{2}$$

für alle $x_0 < x_1 < \cdots < x_{m-1} < x_m < x_{m+1} < \cdots < x_n$ mit $x_n - x_0 < d$ gilt.

Nach dieser Definition ist ohne weiteres ersichtlich, daß insbesondere die in (I.8.1–6) angegebenen Formeln auch für stetige NP gelten. Nach (I.8.2) ist demnach

$$P(A_\varkappa^x \mid A_\iota^w A_\mu^z)\, P(A_\lambda^y \mid A_\varkappa^x A_\mu^z) = P(A_\varkappa^x \mid A_\iota^w A_\lambda^y)\, P(A_\lambda^y \mid A_\iota^w A_\mu^z) \tag{3}$$

für $w < x < y < z$, $z - w < d$ und $\iota, \varkappa, \lambda, \mu = 0, 1$.

Zu diesen Gleichungen ist noch ein Wort zu sagen, jedoch sollen zuerst noch folgende Abkürzungen eingeführt werden.

$$(4) \qquad P(A_1^y | A_\iota^x A_\varkappa^z) = p_{\iota\varkappa}(y|x,z) = 1 - \bar{p}_{\iota\varkappa}(y|x,z) \qquad \iota, \varkappa = \begin{matrix} 0 \\ 1 \end{matrix},$$

falls $x \leqq y \leqq z$ und $z - x < d$ ist. Ferner sei

$$(5) \qquad \frac{p_{\iota\varkappa}(y|x,z)}{\bar{p}_{\iota\varkappa}(y|x,z)} = q_{\iota\varkappa}(y|x,z) = \frac{1}{\bar{q}_{\iota\varkappa}(y|x,z)}.$$

Und wegen (1) gilt

$$(6) \qquad p_{\iota\varkappa}(y+d|x+d, z+d) = p_{\iota\varkappa}(y|x,z).$$

Stillschweigend wird für alles Folgende vorausgesetzt, daß alle in einer Beziehung auftretenden Stellen demselben offenen Intervall der Länge d angehören.

Wie in Kapitel I, § 8 erhält man aus (3) die Gleichungen

$$(7a) \qquad q_{\iota\varkappa}(x|w,z) = q_{\iota 1}(x|w,y)\,\frac{p_{0\varkappa}(y|x,z)}{p_{1\varkappa}(y|x,z)},$$

$$(7b) \qquad q_{\iota\varkappa}(x|w,z) = q_{\iota 0}(x|w,y)\,\frac{\bar{p}_{0\varkappa}(y|x,z)}{\bar{p}_{1\varkappa}(y|x,z)},$$

$$(7c) \qquad q_{\iota\varkappa}(y|w,z) = \frac{p_{\iota 0}(x|w,y)}{p_{\iota 1}(x|w,y)}\,q_{1\varkappa}(y|x,z),$$

$$(7d) \qquad q_{\iota\varkappa}(y|w,z) = \frac{\bar{p}_{\iota 0}(x|w,y)}{\bar{p}_{\iota 1}(x|w,y)}\,q_{0\varkappa}(y|x,z).$$

Diese 16 Gleichungen sind ebenso wie die Gln. (3) hier als ein System von Funktionalgleichungen für die vier Funktionen $p_{\iota\varkappa}(y|x,z)$ bzw. $q_{\iota\varkappa}(y|x,z)$ aufzufassen. Die Lösungen von (3) zu finden, ist das Hauptziel dieses Kapitels. Denn das System von Funktionalgleichungen (3) charakterisiert die stetigen NP in folgendem Sinn:

Sind die Funktionen $p_{\iota\varkappa}(y|x,z)$ Verkettungs-Wn eines stetigen NP, so sind sie notwendig Lösungen von (3). *Umgekehrt läßt sich aus jeder Lösung von* (3) *vermöge der zu* (I.1.1) *analogen Formeln ein stetiger NP konstruieren.*

Dies gilt auch für die Gln. (7, a—d), denn wie in Kapitel I, § 8 angedeutet wurde, sind die Gln. (7, a—d) und (3) gleichwertig. Zu erwähnen ist noch, daß diese 16 Gleichungen, wie zu erwarten, nicht unabhängig voneinander sind. So folgt etwa aus jedem der Quadrupel (7, a—d)

$$\det q_{\iota\varkappa}(y|x,z) = 0.$$

Hierauf sei jedoch nicht näher eingegangen, zumal dieses System von Funktionalgleichungen in § 3 auf ein einfacher gebautes zurückgeführt wird.

Um weitergehende Aussagen zu ermöglichen, wird nun für alles Folgende Stetigkeit und Differenzierbarkeit der Funktionen $p_{\iota\varkappa}(y|x,z)$ im folgenden Umfang vorausgesetzt.

Annahme 1. Die Funktionen $p_{\iota\varkappa}(y|x,z)$ seien für alle $x \leqq y \leqq z$, $z - x < d$ erklärt und dort, abgesehen von $x = y = z$, nebst ihren zweiten partiellen

Ableitungen stetig. Ferner sei

$$(8)\qquad \begin{cases} p_{\iota\varkappa}(t|t,z) = \begin{matrix}0\\1\end{matrix} \text{ für } \begin{matrix}\iota=0\\\iota=1\end{matrix} \text{ und alle } t<z, \\ p_{\iota\varkappa}(t|x,t) = \begin{matrix}0\\1\end{matrix} \text{ für } \begin{matrix}\varkappa=0\\\varkappa=1\end{matrix} \text{ und alle } x<t, \\ p_{\iota\iota}(t|t,t) = \begin{matrix}0\\1\end{matrix} \text{ für } \begin{matrix}\iota=0\\\iota=1\end{matrix} \text{ und alle } t. \end{cases}$$

Hierzu vier Bemerkungen:

1) Man stellt leicht fest, daß diese Annahme den Gln. (3) oder (7, a—d) nicht widerspricht.

2) Ohne Beweis sei erwähnt, daß aus der Existenz der Grenzwerte

$$\lim_{\substack{x\to t\\ y\to t}} p_{\iota\varkappa}(y|x,z) = \begin{matrix}0\\1\end{matrix} \text{ für } \iota = \begin{matrix}0\\1\end{matrix} \text{ und beliebige } t<z$$

$$\lim_{\substack{y\to t\\ z\to t}} p_{\iota\varkappa}(y|x,z) = \begin{matrix}0\\1\end{matrix} \text{ für } \varkappa = \begin{matrix}0\\1\end{matrix} \text{ und beliebige } x<t$$

gleichmäßig in z bzw. x, die Stetigkeit der Funktionen $p_{\iota\varkappa}(y|x,z)$ für alle $x<y<z$ folgt.

Ähnliches gilt für die Differenzierbarkeit.

3) Diese Annahme hat insbesondere zur Folge, daß für $y\to x$

$$P(A_\iota^y)\to P(A_\iota^x) \quad \text{und} \quad P(A_\iota^x A_\varkappa^y)\to \begin{matrix}P(A_\iota^x)\\0\end{matrix} \text{ für } \begin{matrix}\iota=\varkappa\\\iota\neq\varkappa\end{matrix}$$

strebt. Dies folgt aus $P(A_\iota^x A_\varkappa^y) = \sum_\lambda P(A_\varkappa^y|A_\iota^x A_\lambda^x)\ P(A_\iota^x A_\lambda^x)$ für $y\to x$.

4) Aus (3) und (8) folgt noch leicht, daß die Funktionen $p_{\iota\iota}(y|x,z)$ auch noch für $x=y=z$ einseitig stetig sind und dort einseitige partielle Ableitungen zweiter Ordnung besitzen. Dagegen sind die Funktionen $p_{\iota\varkappa}(y|x,z)$ $(\iota\neq\varkappa)$ dort unstetig.

Annahme 2. a) Für $x<y<z$ sei $p_{\iota\varkappa}(y|x,z)\neq 0$ und $\neq 1$.

b) Für $x=y<z$ und $x<y=z$ sei $\frac{\partial}{\partial y} p_{\iota\varkappa}(y|x,z)\neq 0$.

Auch hierzu zwei Bemerkungen:

1) Die Annahme a) ist nicht sehr einschränkend. Denn es kann gezeigt werden, daß $p_{01}(y|x,z)$ und $p_{10}(y|x,z)$ für $x<y<z$ stets von Null und eins verschieden sind. Außerdem folgt aus $p_{11}(y_0|x_0,z_0)=1$ daß für alle $x\leqq y\leqq z$ $p_{11}(y|x,z)=1$ und $p_{00}(y|x,z)=0$ ist. Überdies kann aus der Formel (5) des nächsten Paragraphen abgelesen werden, daß mit der Annahme b) auch a) erfüllt ist.

2) Die Annahme b) wurde getroffen, um zu verhindern, daß der NP in einen oder mehrere offene NP (s. § 5) zerfällt. Es würde hier zu weit führen, auf diesen Fall näher einzugehen.

§ 2. Produktdarstellung der Funktionen $q_{\iota\varkappa}(y|x,z)$

Für $y=y_0$ kann (1.7a) in der Form

$$q_{\iota\varkappa}(x|w,z)=q_{\iota 1}(x|w,y_0)\,f_{\varkappa}(y_0,x,z) \tag{1}$$

mit $w<x<y_0<z$ geschrieben werden. Das heißt, die Funktionen $q_{\iota\varkappa}(x|w,z)$ lassen sich als Produkt zweier Funktionen mit je zwei Veränderlichen darstellen:

$$q_{\iota\varkappa}(x|w,z)=G_{\iota}(x,w)\,H_{\varkappa}(x,z). \tag{2}$$

Die Funktionen G_{ι} und $H_{\varkappa}$ sind jedoch von der gewählten Stelle y_0 abhängig. Dies ist insofern wesentlich, als für festes y_0 die Darstellung (2) nur für $w<x<y_0<z$ gilt. Es ist naheliegend, diese Einschränkung dadurch zu beseitigen, daß man (1) für den Grenzwert $y_0\to x$ betrachtet.

Nach (1.7a) ist für $y=x+h<z$

$$q_{\iota\varkappa}(x|w,z)=q_{\iota 1}(x|w,x+h)\,\frac{p_{0\varkappa}(x+h|x,z)}{p_{1\varkappa}(x+h|x,z)}. \tag{3}$$

Für $h\to 0$ strebt der linke Faktor gegen ∞, und es ist nach (1.5) und (1.8)

$$q_{\iota 1}(x|w,x+h)=\frac{1}{\bar{q}_{\iota 1}(x|w,x+h)}=\frac{1}{h\cdot{}_3\bar{q}_{\iota 1}(x|w,x)+o(h)}, \tag{4}$$

wenn ${}_3\bar{q}_{\iota 1}$ die partielle Ableitung von $\bar{q}_{\iota 1}$ nach der 3. Variablen bedeutet. Allgemein sei unter ${}_nf_{\iota\varkappa}(y|x,z)$ die partielle Ableitung von $f_{\iota\varkappa}$ nach der n-ten Variablen ($n=1,2,3$) verstanden. Wegen $\bar{q}_{\iota 1}(x|w,x)=0$ kann in (4) noch ${}_3\bar{q}_{\iota 1}(x|w,x)$ durch $-{}_1\bar{q}_{\iota 1}(x|w,x)$ ersetzt werden.

Im rechten Faktor von (3) strebt der Nenner für $h\to 0$ gegen 1, der Zähler gegen 0 wie $h\cdot{}_1p_{0\varkappa}(x|x,z)+o(h)=h\cdot{}_1q_{0\varkappa}(x|x,z)+o(h)$. Denn aus (1.5) folgt für $p_{\iota\varkappa}(y|x,z)=0$ $\quad {}_np_{\iota\varkappa}(y|x,z)={}_nq_{\iota\varkappa}(y|x,z)$.

Damit erhält man aus (3) für $h\to 0$

$$q_{\iota\varkappa}(x|w,z)=\frac{-1}{{}_1\bar{q}_{\iota 1}(x|w,x)}\cdot{}_1q_{0\varkappa}(x|x,z). \tag{5}$$

Diese Produktdarstellung gilt für alle $w<x<z$, ist aber bezüglich w und z und ι und $\varkappa$ nicht symmetrisch. Um dies vollends zu erreichen, wird noch die zu (1.7a) symmetrische Formel (1.7c) betrachtet, aus der für $x=y-k>w$ und $k\to 0$

$$q_{\iota\varkappa}(y|w,z)=-{}_1q_{\iota 0}(y|w,y)\cdot\frac{1}{{}_1q_{1\varkappa}(y|y,z)} \tag{6}$$

folgt. Aus (5) und (6) erhält man schließlich bei passender Änderung der Argumente die symmetrische Produktdarstellung

$$q_{\iota\varkappa}(y|x,z)=\sqrt{\frac{{}_1q_{\iota 0}(y|x,y)}{{}_1\bar{q}_{\iota 1}(y|x,y)}}\cdot\sqrt{\frac{{}_1q_{0\varkappa}(y|y,z)}{{}_1\bar{q}_{1\varkappa}(y|y,z)}} \tag{7}$$

Nach (5), (6) oder (7) gibt es also vier Funktionen $K_\iota(y, x)$ mit $y > x$ und $L_\iota(y, z)$ mit $y < z$, $\iota = 0, 1$, so daß

$$q_{\iota\varkappa}(y|x, z) = K_\iota(y, x)\, L_\varkappa(y, z) \tag{8}$$

für alle $x < y < z$ ist. Setzt man noch für beliebige Werte von y

$$K_0(y, y) = L_0(y, y) = 0, \quad K_1(y, y) = L_1(y, y) = \infty, \tag{9}$$

so gilt (8) nach Annahme 1 für alle $x \leqq y \leqq z$. (Wegen des Symbols ∞ beachte man das in Kapitel I, § 3 Gesagte.) Um die Unstetigkeit von K_1 und L_1 bei $y = z$ zu umgehen, werden neben K_ι und L_ι noch die Funktionen

$$M_\iota(y, x) = K_\iota^{1-2\iota}(y, x) \quad \text{und} \quad N_\iota(y, z) = L_\iota^{1-2\iota}(y, z) \qquad \iota = 0, 1 \tag{10}$$

betrachtet. Dabei ist für alle y

$$M_\iota(y, y) = N_\iota(y, y) = 0, \qquad \iota = 0, 1. \tag{11}$$

Für alles Folgende werden nun

$$K_\iota(y, x) = + \sqrt{\frac{{}_1q_{\iota 0}(y|x, y)}{{}_1\bar{q}_{\iota 1}(y|x, y)}}, \qquad L_\iota(y, z) = + \sqrt{\frac{{}_1\bar{q}_{0\iota}(y|y, z)}{{}_1\bar{q}_{1\iota}(y|y, z)}} \tag{12}$$

für $\iota = 0, 1$ gesetzt. Wegen Annahme 1 und 2 sind die so gewählten Funktionen K_ι und L_ι, sowie M_ι und N_ι für alle $x \neq y$ bzw. $y \neq z$ von Null verschieden und besitzen dort stetige Ableitungen. Aus (8), (10) und (11) folgt, daß M_ι und N_ι für $x = y$ bzw. $y = z$ noch einseitige, stetige und von Null verschiedene Ableitungen haben. Die zweiten partiellen Ableitungen von M_ι und N_ι existieren für $x = y$ bzw. $y = z$ ebenfalls und sind dort stetig, was für $M_0(y, x)$ gezeigt sei.

Aus (8) und (10) folgt

$$q_{00}(y|x, z) = M_0(y, x)\, N_0(y, z), \tag{13}$$

$${}_1q_{00}(y|x, z) = M_{0y}(y, x)\, N_0(y, z) + M_0(y, x)\, N_{0y}(y, z). \tag{14}$$

Dabei bedeute $M_{\iota y}$, $N_{\iota y}$ bzw. $M_{\iota x}$, $N_{\iota x}$ stets die partielle Ableitung von M_ι, N_ι nach der 1. bzw. 2. Variablen. Ersetzt man in (14) einmal x, y durch y, $y + k$ und einmal durch y, y und bildet den Differenzenquotienten, so ergibt sich

$$\begin{aligned}\tfrac{1}{k}\,[{}_1q_{00}(y + k|y, z) - {}_1q_{00}(y|y, z)] = \tfrac{1}{k}\,[&M_{0y}(y + k, y)\, N_0(y + k, z) - \\ &- M_{0y}(y, y)\, N_0(y, z) + M_0(y + k, y)\, N_{0y}(y + k, z)].\end{aligned}$$

Nach Annahme 1 existiert die linke Seite dieser Gleichung für $k \to 0$, und der Grenzwert ist stetig in y. Ferner ist $N_0(y + k, z) = N_0(y, z) + O(k)$. Also ist die rechte Seite gleich

$$\tfrac{1}{k}\,[M_{0y}(y + k, y) - M_{0y}(y, y)]\, N_0(y, z) + \tfrac{1}{k}\, O(k).$$

Dabei ist der letzte Summand stetig in y. Wegen $N_0(y, z) \neq 0$ existiert also $M_{0yy}(y, y)$, und der Grenzwert ist stetig. Leicht folgt aus (13) die Existenz

und Stetigkeit von $M_{0\varkappa z}(y, x)$ und aus (14) die von $M_{0\varkappa y}(y, x)$. Setzt man schließlich (8) und (10) in (12) ein, so erhält man noch die Beziehungen

$$N_{0\varkappa}(y, y) = N_{1\varkappa}(y, y) \quad \text{und} \quad M_{0\varkappa}(y, y) = M_{1\varkappa}(y, y). \tag{15}$$

§ 3. Die Funktionalgleichungen für $K_\iota(y, x)$ und $L_\iota(y, z)$ und das Verhalten der Funktionen $q_{\iota\varkappa}(y|y-h, y+h)$ für kleine h

Das Problem der Auflösung des Systems von Funktionalgleichungen (1.7, a—d) läßt sich, wenn man nur Funktionen $p_{\iota\varkappa}(y|x, z)$ betrachtet, die den Annahmen 1 und 2 genügen, auf die Lösung von partiellen Differentialgleichungen zurückführen. In diesen treten noch willkürliche Funktionen auf, die sich aus dem Verhalten von $q_{\iota\varkappa}(x|x-h, x+h)$ für kleine h bestimmen lassen. Es wird sich zeigen, daß $q_{\iota\varkappa}(y|x, z)$ durch Vorgabe von $q_{\iota\varkappa}(x|x-h, x+h)$ für $h \to 0$ eindeutig bestimmt ist.

Hierbei ist wesentlich, daß mit Hilfe der eben abgeleiteten Produktdarstellung das Problem auf die Aufstellung bzw. Lösung von Differentialgleichungen für K_ι und L_ι zurückgeführt werden kann. Dazu sollen zunächst die Funktionalgleichungen (1.7, a—d) auf die Funktionen K_ι und L_ι umgeschrieben werden. Man beachte, daß hierdurch sowohl die Zahl der Gleichungen als auch die Zahl der unabhängigen Variablen reduziert wird.

Aus (1.7a), (1.5) und (2.8) folgt zunächst

$$\begin{cases} K_\iota(x, w)\, L_\varkappa(x, z) = K_\iota(x, w)\, L_1(x, y)\, \dfrac{1 + K_1^{-1}(y, x)\, L_\varkappa^{-1}(y, z)}{1 + K_0^{-1}(y, x)\, L_\varkappa^{-1}(y, z)} \\ L_\varkappa(x, z) = L_1(x, y)\, \dfrac{K_0(y, x)}{K_1(y, x)}\, \dfrac{1 + K_1(y, x)\, L_\varkappa(y, z)}{1 + K_0(y, x)\, L_\varkappa(y, z)}. \end{cases} \tag{1}$$

Für $\varkappa = 0$ und $z \to y$ folgt hieraus und aus (2.9)

$$\begin{cases} K_0(y, x)\, L_1(x, y) = K_1(y, x)\, L_0(x, y) \\ \text{oder auch} \\ M_0(y, x)\, M_1(y, x) = N_0(x, y)\, N_1(x, y). \end{cases} \tag{2}$$

Aus (1.7c) erhält man ebenso

$$K_\iota(y, w) = \frac{1 + K_\iota(x, w)\, L_1(x, y)}{1 + K_\iota(x, w)\, L_0(x, y)}\, \frac{L_0(x, y)}{L_1(x, y)}\, K_1(y, x). \tag{3}$$

Beachtet man noch Gl. (2), die übrigens auch aus (3) folgt, so erhält man aus (1) und (3) die vier Funktionalgleichungen

$$K_\iota(z, x) = K_0(z, y)\, \frac{1 + L_1(y, z)\, K_\iota(y, x)}{1 + L_0(y, z)\, K_\iota(y, x)}, \tag{4}$$

$$L_\iota(x, z) = L_0(x, y)\, \frac{1 + K_1(y, x)\, L_\iota(y, z)}{1 + K_0(y, x)\, L_\iota(y, z)} \tag{5}$$

für alle $x < y < z$.

Diese Gleichungen, die ein Analogon zu den Chapman-Kolmogoroff-Gleichungen bilden, besagen unter anderem, daß sich die Funktionswerte von

K_ι und L_ι an der Stelle x, z aus denen von K_ι und L_ι an den Stellen x, y und y, z berechnen lassen. Wie man sich leicht überzeugt, erhält man aus (1.7b) und (1.7d) ebenfalls die Gln. (4) und (5).

Nun soll das Verhalten von $q_{\iota\varkappa}(x|x-h, x+h)$ für $h \to 0$ untersucht werden. Dazu zieht man am besten die Faktoren $M_\iota(y, x)$ und $N_\iota(x, y)$ heran. Nach § 2 sind diese Funktionen für $x=y$ gleich Null und besitzen dort stetige partielle Ableitungen zweiter Ordnung. Also gibt es vier (auch für $x=y$) stetig differenzierbare Funktionen $f_\iota(y, x)$ und $g_\iota(x, y)$ mit

$$M_\iota(y, x) = (y-x) f_\iota(y, x) \quad \text{und} \quad N_\iota(x, y) = (y-x) g_\iota(x, y).$$

Für $h, k \geqq 0$ gilt hiernach

$$M_\iota(y+k, y-h) = (h+k)\left[f_\iota(y, y) - h f_{\iota x}(y, y) + k f_{\iota y}(y, y)\right] + o(h^2+k^2).$$

Beachtet man noch, daß $\frac{d}{dy} f_\iota(y, y) = f_{\iota x}(y, y) + f_{\iota y}(y, y)$ ist, so erhält man die Darstellung

$$(6)\quad M_\iota(y+k, y-h) = (h+k)\left[m_\iota(y) + k r_\iota(y) - h\left(m_\iota'(y) - r_\iota(y)\right)\right] + o(h^2+k^2),$$

wobei $m_\iota'(y)$ und $r_\iota(y)$ stetige Funktionen sind. Ferner ist $m_\iota(y) = M_{\iota y}(y, y) = -M_{\iota x}(y, y) > 0$. Ebenso gilt

$$(7)\quad N_\iota(y-h, y+k) = (h+k)\left[n_\iota(y) + k s_\iota(y) - h\left(n_\iota'(y) - s_\iota(y)\right)\right] + o(h^2+k^2)$$

mit $n_\iota(y) = N_{\iota x}(y, y) = -N_{\iota y}(y, y)$.

Dabei ist nach (2.13) $m_0(y) = m_1(y)$ und $n_0(y) = n_1(y)$. Hieraus folgt, wenn man (6) und (7) in (2) einsetzt und $h, k \to 0$ streben läßt, $m_0^2 = n_0^2$, oder wegen $m_0, n_0 > 0$

$$m_0 = n_0 = m, \quad \text{sowie} \quad r_0 + r_1 = s_0 + s_1.$$

Damit erhält man für $\iota = 0, 1$

$$(8)\quad \begin{cases} M_\iota(y+k, y-h) = (h+k)\left[m(y) + k r_\iota(y) + h\left(r_\iota(y) - m'(y)\right)\right] + o(h^2+k^2) \\ N_\iota(y-h, y+k) = (h+k)\left[m(y) + k s_\iota(y) + h\left(s_\iota(y) - m'(y)\right)\right] + o(h^2+k^2) \end{cases}$$

mit

$$r_0(y) + r_1(y) = s_0(y) + s_1(y).$$

Ferner strebt für $h \to 0$

$$q_{00}(y|y-h, y+h) = M_0(y, y-h) \cdot N_0(y, y+h) \to m^2 h^2$$

$$q_{01}(y|y-h, y+h) = \frac{M_0(y, y-h)}{N_1(y, y+h)} \to \frac{m + (r_0 - m')h}{m + s_1 h} \to 1 + \frac{r_0 - s_1 - m'}{m} h = 1 + \gamma h$$

$$q_{10}(y|y-h, y+h) = \frac{N_0(y, y+h)}{M_1(y, y-h)} \to \frac{m + s_0 h}{m + (r_1 - m')h} \to 1 - \frac{r_1 - s_0 - m'}{m} h = 1 - \delta h$$

$$q_{11}(y|y-h, y+h) = M_1(y, y-h) \cdot N_1(y, y+h) \to m^2 h^2.$$

Dabei ist

$$\gamma = \gamma(y) = \frac{r_0 - s_1 - m'}{m}, \quad \delta = \delta(y) = \frac{r_1 - s_0 - m'}{m} \quad \text{und} \quad \gamma + \delta = -2\frac{m'}{m}.$$

Setzt man noch

$$\frac{r_0 - s_1}{m} = \frac{s_0 - r_1}{m} = n = n(y), \tag{9}$$

so erhält man zusammenfassend für $h \to 0$

$$q_{\iota\varkappa}(y|y-h, y+h) \to \begin{cases} m^2(y)\,h^2 \\ 1+\left(n(y) - \dfrac{m'(y)}{m(y)}\right)h \\ 1+\left(n(y) + \dfrac{m'(y)}{m(y)}\right)h \\ \dfrac{1}{m^2(y)\,h^2} \end{cases} \quad \text{für } \iota,\varkappa = \begin{cases} 0,0 \\ 0,1 \\ 1,0 \\ 1,1. \end{cases} \tag{10}$$

§ 4. Die Differentialgleichungen für $K_\iota(y, x)$ und $L_\iota(y, z)$

Für $z = y + k$ folgt aus (3.4)

$$K_\iota(y+k, x) - K_\iota(y, x) = K_0(y+k, y)\,\frac{1 + L_1(y, y+k)\,K_\iota(y, x)}{1 + L_0(y, y+k)\,K_\iota(y, x)} - K_\iota(y, x).$$

Dividiert man beide Seiten mit k, so strebt die linke Seite für $k \to 0$ gegen $K_{\iota,y}(y, x)$, und rechts ergibt sich

$$\frac{1}{k}\,\frac{K_0(y+k, y) + (K_0(y+k, y)\,L_1(y, y+k) - 1)\,K_\iota(y, x) - L_0(y, y+k)\,K_\iota^2(y, x)}{1 + L_0(y, y+k)\,K_\iota(y, x)}.$$

Dies ist nach (3.8) und (2.10) gleich

$$m(y) + \frac{K_0(y+k, y)\,L_1(y, y+k) - 1}{k}\,K_\iota(y, x) - m(y)\,K_\iota^2(y, x) + \frac{1}{k}\,o(k).$$

Schließlich strebt für $k \to 0$

$$\frac{1}{k}\left(K_0(y+k, y)\,L_1(y, y+k) - 1\right) \to \frac{1}{k}\left(\frac{m + r_0 k}{m + s_1 k} - 1\right) \to \frac{r_0 - s_1}{m} = n(y).$$

Insgesamt erhält man für $k \to 0$ die Differentialgleichung

$$K_{\iota,y}(y, x) = m(y) + n(y)\,K_\iota(y, x) - m(y)\,K_\iota^2(y, x), \qquad \iota = 0, 1. \tag{1}$$

Ebenso folgt aus (3.5) für $x = y - h$

$$L_{\iota,y}(y, z) = -\,m(y) - n(y)\,L_\iota(y, z) + m(y)\,L_\iota^2(y, z), \qquad \iota = 0, 1. \tag{2}$$

(1) und (2) sind die gesuchten partiellen Differentialgleichungen für die Funktionen $K_\iota(y, x)$ und $L_\iota(y, z)$. Sie entsprechen den vorwärts- und rückwärtsgenommenen Differentialgleichungen der Markoff-Prozesse[8]). Die Anfangsbedingungen von (1) und (2) sind

$$K_0(x, x) = 0, \quad K_1(x, x) = \infty; \quad L_0(z, z) = 0, \quad L_1(z, z) = \infty. \tag{3}$$

Zu beachten ist noch, daß die Differentialgleichungen nicht gekoppelt sind. Und da nur Ableitungen nach einer Variablen auftreten, können sie wie gewöhnliche Differentialgleichungen vom Riccati-Typ behandelt werden.

[8]) Siehe etwa W. Feller [3], S. 426.

Außerdem tritt noch folgende Vereinfachung ein. Ist nämlich $f(y)$ eine Lösung von (1), so ist $-\frac{1}{f(y)}$ eine Lösung von (2). Denn aus $f' = m + nf - mf^2$ folgt durch Division mit f^2

$$\left(-\frac{1}{f}\right)' = m\left(-\frac{1}{f}\right)^2 - n\left(-\frac{1}{f}\right) - m.$$

Die Funktionen $K_0(y, x)$, $K_1(y, x)$, $-L_0^{-1}(y, z)$, $-L_1^{-1}(y, z)$ genügen also alle der Differentialgleichung (1).

Das allgemeine Integral von (1) ist von der Form

$$F(y, x) = \frac{a(y) + C(x)\, b(y)}{c(y) + C(x)\, d(y)}, \tag{4}$$

wobei $C(x)$ eine willkürliche Funktion von x ist.

Für $K_0(y, x)$ bestimmt sich $C(x)$ nach (3) aus $0 = a(y) + C(x)\, b(y)$. Es ist also

$$K_0(y, x) = \frac{a(y)\, b(x) - a(x)\, b(y)}{c(y)\, b(x) - a(x)\, d(y)}. \tag{5}$$

$K_1(y, x)$ läßt sich nach (1) ebenfalls auf die Form (4) mit denselben Funktionen $a(y), \ldots, d(y)$ bringen. Die Anfangsbedingung (3) ergibt hier $0 = c(x) + C(x)\, d(x)$ und damit

$$K_1(y, x) = \frac{a(y)\, d(x) - c(x)\, b(y)}{c(y)\, d(x) - c(x)\, d(y)}. \tag{6}$$

$-L_i^{-1}(y, z)$ ist ebenfalls von der Gestalt (4), wenn man dort x durch z ersetzt. Unter Beachtung der Anfangsbedingungen erhält man danach

$$L_0(y, z) = -\frac{c(y)\, d(z) - c(z)\, d(y)}{a(y)\, d(z) - c(z)\, b(y)}, \tag{7}$$

$$L_1(y, z) = -\frac{c(y)\, b(z) - a(z)\, d(y)}{a(y)\, b(z) - a(z)\, b(y)}. \tag{8}$$

Die Funktionen $K_i(y, x)$ und $L_i(y, z)$ sind also durch Vorgabe der Funktionen $m(y)$ und $n(y)$, sowie der Anfangsbedingungen (3) eindeutig bestimmt. Durch (5) und (6) ist $K_i(y, x)$ übrigens auch für $x \geqq y$ erklärt, und zwar ist nach (7) und (8)

$$K_i(y, z) = -L_{1-i}^{-1}(y, z). \tag{9}$$

Ferner ist mit $m(y + d) = m(y)$ und $n(y + d) = n(y)$ auch

$$K_i(y + d, x + d) = K_i(y, x).$$

Denn dann genügt $R_i(y, x) = K_i(y + d, x + d)$ ebenfalls der Differentialgleichung (1) mit den Anfangsbedingungen $R_i(x, x) = 0, \infty$ für $i = 0, 1$. Dasselbe gilt für $L_i(y, z)$.

Zu zeigen ist jetzt noch, daß die Lösungen von (1) und (2), $K_i(y, x)$ und $L_i(y, z)$ auch den Funktionalgleichungen (3.4) und (3.5) und damit auch (1.3)

genügen. Betrachtet man $K_0(z, x)$ und $K_0(y, x)$ bei festem y und z als Funktionen von x, so besteht nach (3.4) zwischen diesen Funktionen der Zusammenhang

$$K_0(z, x) = R\,\frac{1 + S\,K_0(y, x)}{1 + T\,K_0(y, x)}\,. \tag{10}$$

Setzt man (5) in (10) ein, so müssen sich daraus, wenn (5) die Funktionalgleichung (3.4) befriedigt, R, S und T unabhängig von x bestimmen lassen. Dies ist in der Tat der Fall, und zwar erhält man für R, S und T, wie gefordert, die rechten Seiten von (5), (7) und (10), wenn man in (5) noch x, y durch y, z ersetzt. Dasselbe gilt für $K_1(y, x)$ und $L_i(y, z)$.

Bemerkenswert ist noch, daß demnach (5) bis (8) für beliebige Funktionen $a(x), \ldots, d(x)$ eine Lösung von (3.4) und (3.5) sind. Allerdings ist dadurch Gl. (2.15) im allgemeinen nicht erfüllt.

Zusammenfassend kann man nun sagen:

Durch Vorgabe von

$$\lim_{h \to 0} h^2 q_{00}(y|y - h, y + h) = m^2(y)\,,$$

$$\lim_{h \to 0} \frac{1}{h}\left(q_{01}(y|y - h, y + h) - 1\right) = \gamma(y)\,,$$

wobei $m(y)$ eine stetig differenzierbare Funktion > 0 und $\gamma(y)$ eine stetige Funktion ist, sind die Funktionen $q_{ik}(y|x, z)$ eines NP durch die Lösungen der Differentialgleichungen (1), (2) *mit* $n = \frac{m'}{m} + \gamma$ *und den Anfangsbedingungen* (3) *eindeutig bestimmt.*

Ist der NP homogen, d.h. ist $q_{ik}(y|x, z)$ nur eine Funktion von $y - x$ und $z - y$ oder $q_{ik}(y|y - s, y + t) = f_{ik}(s, t)$, so läßt sich die Differentialgleichung (1) geschlossen integrieren. In diesem Fall ist nämlich nach (3.10) $m(y) = m = \text{const}$ und $n(y) = n = \text{const}$, und die Differentialgleichung lautet

$$f'(y) = m + n f(y) - m f^2(y).$$

Ihre allgemeine Lösung ist

$$f(y) = \frac{1}{a}\,\frac{a^2 - C e^{-r y}}{1 + C e^{-r y}}$$

mit $r = \sqrt{n^2 + 4m^2}$, $a = (n + r) : 2m$ und der Integrationskonstanten C. Aus den Anfangsbedingungen für $K_i(y, x)$ und (9) erhält man hieraus

$$K_0(y, x) = a\,\frac{e^{r\cdot(y-x)} - 1}{e^{r\cdot(y-x)} + a^2}\,, \qquad K_1(y, x) = \frac{1}{a}\,\frac{a^2 e^{r\cdot(y-x)} + 1}{e^{r\cdot(y-x)} - 1}$$

$$L_0(y, z) = a\,\frac{e^{r\cdot(z-y)} - 1}{e^{r\cdot(z-y)} + a^2}\,, \qquad L_1(y, z) = \frac{1}{a}\,\frac{a^2 e^{r\cdot(z-y)} + 1}{e^{r\cdot(z-y)} - 1}$$

und damit

$$q_{00}(y|x, z) = a^2\,\frac{e^{r\cdot(y-x)} - 1}{e^{r\cdot(y-x)} + a^2}\,\frac{e^{r\cdot(z-y)} - 1}{e^{r\cdot(z-y)} + a^2} \text{ usw.}$$

Vermerkt sei noch, daß man dieses Ergebnis auch durch einen Grenzübergang aus den Formeln (I.8.12) erhält.

§ 5. Die Randwerte und die Berechnung der absoluten Wn

Die bisherigen Ableitungen gelten insofern sowohl für Ereignisse A_i^x, die einer geschlossenen, als auch für solche, die einer offenen Kurve zugeordnet sind, als durch $m(y)$, $n(y)$, gegeben für $a \leqq y \leqq b$, $q_{ik}(y|x, z)$ für alle $a \leqq x \leqq y \leqq z \leqq b$ bestimmt ist. Dieser Unterschied ist aber wesentlich bei der Berechnung der absoluten und der Übergangs-Wn.

Ist die Kurve geschlossen, so sind nach Voraussetzung $m(y)$ und $n(y)$ d-periodische Funktionen. Der NP heißt in diesem Fall geschlossen, wenn noch $m(y) > 0$ ist.

Der stetige NP heißt offen, wenn A_i^x auf einer offenen Kurve $(0 \leqq x \leqq c)$ erklärt ist, und die bedingten Wn die Forderungen (1.2) für alle $0 \leqq x_0 < \cdots < x_n \leqq c$ und

$$P(A^0|A^x A^c) = P(A^0|A^x) \quad \text{für } 0 < x < c \tag{1}$$

erfüllen.

Man beachte, daß in diesem Fall die Forderung $A_i^0 = A_i^c$ nicht besteht.

Für beide Fälle wird nun noch das Verhalten am Rande und die Berechnung der absoluten Wn $P(A_1^x)$ angegeben. Dabei werden die Abkürzungen

$$q_i(y|x) = Q(A_1^y|A_i^x) \quad \text{und} \quad q(x) = Q(A_1^x)$$

benützt.

1) Der offene NP. Auch hier sei vorausgesetzt, daß die Funktionen $p_{ik}(y|x, z)$ für $0 \leqq x \leqq y \leqq z \leqq c$ den Annahmen 1 und 2 genügen. Dies bedeutet insbesondere, daß die Funktionen $m(y)$, $n(y)$ in $0 < y < c$ erklärt sind, $m(y)$ dort > 0 ist, und die Grenzwerte von $m(y)$ und $n(y)$ für $y \to 0$ und $y \to c$ existieren. Sie seien mit $m(0)$, $n(0)$, $m(c)$ und $n(c)$ bezeichnet.

Durch Vorgabe der Funktionen $m(y)$, $n(y)$ für $0 \leqq y \leqq c$ können also die Wv $q_{ik}(y|x, z)$ für $0 \leqq x \leqq y \leqq z \leqq c$ durch Auflösen der Differentialgleichungen von § 4 berechnet werden. Die absoluten Wn $P(A_i^x)$, sowie die Übergangs-Wn $P(A_k^y|A_i^x)$ sind hierdurch jedoch nicht bestimmt.

Nach (1) und (I.5.5–7) bestehen zwischen den Wv $q_i(y|x)$ und $q_i(x|y)$ die Zusammenhänge

$$\frac{q_1(y|x)}{q_0(y|x)} = \frac{q_1(x|y)}{q_0(x|y)} = \frac{K_1(y, x)}{K_0(y, x)} = \frac{L_1(x, y)}{L_0(x, y)}, \tag{2}$$

$$q_1(y|x) = \frac{1 + q_1(z|y)}{L_0^{-1}(y, z) + L_1^{-1}(y, z)\, q_1(z|y)}\, K_1(y, x), \tag{3}$$

$$q_1(y|z) = \frac{1 + q_1(x|y)}{K_0^{-1}(y, x) + K_1^{-1}(y, x)\, q_1(x|y)}\, L_1(y, z) \tag{4}$$

für $x < y < z$. Setzt man nun noch voraus, daß

$$\lim_{h \to 0} h\, q_1(0|h) = l \quad \text{und} \quad \lim_{h \to 0} h\, q_1(c|c - h) = r \tag{5}$$

existiert und gibt die Randwerte l und r vor, so lassen sich aus (2), (3) und (4) die Wv $q_i(y|x)$ und $q_i(x|y)$, sowie $q(x)$ und damit nach den Methoden von Kapitel I, § 2 alle übrigen absoluten Wv berechnen. Daß die Rand-Wv

$l \geqq 0$ und $r \geqq 0$ beliebig und unabhängig vorgegeben werden können, zeigt man wie beim diskreten offenen NP.

Aus (3) folgt für $y = c - h$, $z = c$

$$q_1(c-h|x) = \frac{1 + q_1(c|c-h)}{L_0^{-1}(c-h,c) + L_1^{-1}(c-h,c)\, q_1(c|c-h)} K_1(c-h,x).$$

Für $h \to 0$ erhält man hieraus nach (5) und (3.8)

$$q_1(c|x) = r\, m(c)\, K_1(c,x). \tag{6}$$

Ebenso folgt aus (4) für $x = 0$, $y = h$ und $h \to 0$

$$q_1(0|z) = l\, m(0)\, L_1(0,z). \tag{7}$$

Insbesondere ist also

$$q_1(c|0) = r\, m(c)\, K_1(c,0), \qquad q_1(0|c) = l\, m(0)\, L_1(0,c),$$

und aus (2) folgt

$$q_0(c|0) = r\, m(c)\, K_0(c,0), \qquad q_0(0|c) = l\, m(0)\, L_0(0,c).$$

Aus (I.4.3) erhält man schließlich noch

$$q(0) = l \cdot m(0)\, L_0(0,c)\, \frac{1 + r\, m(c)\, K_1(c,0)}{1 + r\, m(c)\, K_0(c,0)}.$$

Die Berechnung von $q(x)$ erfolgt ebenso aus den Wv $q_i(y|x)$ und $q_i(x|y)$, wobei man die letzteren durch Einsetzen von (6) und (7) in (3) und (4) bei passender Änderung der Argumente erhält.

2) Der geschlossene NP. In diesem Fall sind durch Vorgabe von $m(y) > 0$ und $n(y)$, wie jetzt gezeigt werden soll, auch die Übergangs-Wv $q_i(y|x)$ für $x \gtrless y$ und damit auch wieder alle absoluten Wn bestimmt. Wendet man (I.5.3) auf $A = A_i^x$, $B = A_1^y$, $C = A_1^z$ $(x < y < z)$ an und beachtet, daß wegen der Einschränkung $z - x < d$ für die Abkürzungen (1.4) $Q(B|AC) = q_{i1}(y|x,z)$ und $Q(C|AB) = q_{1i}(z|y, x+d)$ zu setzen ist, so erhält man

$$q_i(y|x) = K_i(y,x)\, L_1(y,z)\, \frac{K_0(z,y)}{K_1(z,y)}\, \frac{1 + K_1(z,y)\, L_i(z, x+d)}{1 + K_0(z,y)\, L_i(z, x+d)}$$

und daraus nach (3.1)

$$q_i(y|x) = K_i(y,x)\, L_i(y, x+d) \quad \text{für } x < y. \tag{8}$$

Es ist also

$$q_i(y|x) = q_{ii}(y|x, x+d) \quad \text{für } x < y,$$

was formal auch aus $Q(A_1^y | A_i^x A_i^{x+d}) = Q(A_1^y | A_i^x A_i^x) = Q(A_1^y | A_i^x)$ folgt. Genau so erhält man ebenfalls für $x < y$

$$q_i(x|y) = K_i(x, y-d)\, L_i(x,y). \tag{9}$$

Die Beziehungen (8) und (9) treten hier an Stelle der Randbedingungen bei offenen NP.

Wieder aus (I.4.3), sowie aus (8) und (9) erhält man noch

$$q(x) = K_0(x, y-d)\, L_0(x, y)\, \frac{1 + K_1(y, x)\, L_1(y, x+d)}{1 + K_0(y, x)\, L_0(y, x+d)}.$$

Diese Gleichung läßt sich entweder durch Umformung von $L_0(x, y)$ unter Verwendung von (3.5) oder bequemer durch den Grenzübergang $y \to x$ auf die einfache Form

$$q(x) = K_0(x, x-d)\, L_1(x, x+d)$$

bringen.

Für den auf der geschlossenen Kurve $0 \leqq x \leqq d$ durch $m = \text{const}$ und $n = \text{const}$ definierten homogenen geschlossenen NP erhält man hiernach und aus den im letzten Paragraphen am Schluß abgeleiteten Beziehungen

$$q(x) = \frac{a^2 e^{\gamma d} + 1}{e^{\gamma d} + a^2}.$$

$q(x)$ ist also konstant. Dies war zu erwarten; denn durch Vorgabe von $m = \text{const}$ und $n = \text{const}$ ist keine Stelle der geschlossenen Kurve vor der anderen ausgezeichnet.

Literatur

[1] BARTLETT, M.: An Introduction to Stochastic Processes. Cambridge 1955. — [2] DOOB, J.: Stochastic Processes. New York 1953. — [3] FELLER, W.: An Introduction to Probability Theory and its Applications. Vol. I. 2nd ed. New York 1957. — [4] MONTROLL, E.: Statistical mechanics of nearest neighbor systems. J. Chem. Phys. **9**, 706–721 (1941). — [5] MONTROLL, E.: Statistical mechanics of nearest neighbor systems. J. Chem. Phys. **10**, 61–77 (1942).

Mathematisches Institut A der Technischen Hochschule Stuttgart

(Eingegangen am 5. August 1958)

KEOGH, F. R., and G. M. PETERSEN
Math. Zeitschr. 71, 31–35 (1959)

A strengthened form of a theorem of WIENER

By
F.R. KEOGH and G.M. PETERSEN

We begin with two definitions relating to a sequence $\{s_n\}$, $n=0, 1, \ldots$, of complex numbers.

1. $\{s_n\}$ is said to be almost convergent to s if, as $p \to \infty$, $(p+1)^{-1} \sum_{n}^{n+p} s_m$ tends to s uniformly with respect to n [1].

2. $\{s_n\}$ is said to be almost periodic if, corresponding to any $\varepsilon > 0$, there are numbers N, l such that in every interval $(k, k+l)$ $(k \geq 0)$, there is an integer p for which $|s_{n+p} - s_n| < \varepsilon$ for all $n \geq N$ [1].

Suppose that $F(\vartheta)$ is a real function of bounded variation over $(0, 2\pi)$, and let

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} e^{in\vartheta} dF(\vartheta), \qquad n = 0, 1, \ldots. \tag{1}$$

A well known theorem of WIENER [2] may be interpreted as stating that *a necessary and sufficient condition for $F(\vartheta)$ to be continuous is that the sequence $\{|c_n|^2\}$ (or the sequence $\{|c_n|\}$) should be $(C, 1)$ summable to zero.* Suppose that

$$F(\vartheta) = G(\vartheta) + H(\vartheta), \tag{2}$$

where $G(\vartheta) = \sum_{\lambda_m \leq \vartheta} \mu_m$ is a function with discontinuities μ_m at λ_m, $m = 1, 2, \ldots$, and $H(\vartheta)$ is a continuous function. We shall prove the following stronger form of WIENER's theorem:

THEOREM. (i) *$\{|c_n|^2\}$ is almost convergent to $(4\pi)^{-2} \sum_{1}^{\infty} \mu_m^2$.* (ii) *For $F(\vartheta)$ to be continuous it is necessary that $\{|c_n|^2\}$ (or $\{|c_n|\}$) should be almost convergent to zero and sufficient that $\{|c_n|^2\}$ (or $\{|c_n|\}$) should be summable to zero by some summation method which contains almost convergence.*

Since the method of $(C, 1)$ summation contains almost convergence [1], part (ii) of our theorem strengthens the necessary condition and weakens the sufficient condition of WIENER's theorem. We shall require the following lemmas, the proofs of which, if not already in the literature, are elementary and will be omitted.

LEMMA 1. *For any real λ the sequence $\{e^{i\lambda n}\}$ is almost periodic.*

LEMMA 2. *Suppose that for each m, $m = 1, 2, \ldots$, $\{s_m(n)\}$ is almost periodic and that, as $m \to \infty$, $s_m(n)$ tends uniformly to $t(n)$. Then $\{t(n)\}$ is almost periodic.*

LEMMA 3. *If* $\{s_n\}$, $\{t_n\}$ *are almost periodic then so are* $\{|s_n|\}$, $\{s_n t_n\}$ *and* $\{s_n+t_n\}$.

LEMMA 4. *An almost periodic sequence is almost convergent.*

Let $c_n=a_n+b_n$ be the decomposition of c_n corresponding to (2), i.e.

$$a_n=\frac{1}{2\pi}\int_0^{2\pi} e^{in\vartheta}\,dG(\vartheta), \tag{3}$$

$$b_n=\frac{1}{2\pi}\int_0^{2\pi} e^{in\vartheta}\,dH(\vartheta). \tag{4}$$

By squaring the inequalities $||a_m|-|b_m||\leqq|c_m|\leqq|a_m|+|b_m|$, summing from n to $n+p$ and using SCHWARZ's inequality we obtain

$$\begin{aligned}\sum|a_m|^2-2\left(\sum|a_m|^2\right)^{\frac{1}{2}}\left(\sum|b_m|^2\right)^{\frac{1}{2}}+\sum|b_m|^2&\leqq\sum|c_m|^2\\&\leqq\sum|a_m|^2+2\left(\sum|a_m|^2\right)^{\frac{1}{2}}\left(\sum|b_m|^2\right)^{\frac{1}{2}}+\sum|b_m|^2,\end{aligned}$$

all sums being from n to $n+p$. Thus, dividing by $p+1$, to prove part (i) of the theorem it is enough to prove that $\{|a_n|^2\}$ is almost convergent to $(4\pi)^{-2}\sum\mu_m^2$, $\{|b_n|^2\}$ almost convergent to zero. We first prove the latter. From (4),

$$\left\{\begin{aligned}&(p+1)^{-1}\sum_n^{n+p} b_m=\frac{1}{2\pi}\int_0^{2\pi} T(n,p,\vartheta)\,dH(\vartheta),\\&\text{where}\\&\qquad T(n,p,\vartheta)=(p+1)^{-1}e^{in\vartheta}\,\frac{1-e^{i(p+1)\vartheta}}{1-e^{i\vartheta}}.\end{aligned}\right. \tag{5}$$

Since $|T|=(p+1)^{-1}|\sin\frac{1}{2}(p+1)\vartheta|\operatorname{cosec}\frac{1}{2}\vartheta$ we have $|T|\leqq 1$ and, for a given $\varepsilon>0$, $|T|\leqq(p+1)^{-1}\operatorname{cosec}\frac{1}{2}\varepsilon$ in $(\varepsilon,2\pi)$. Thus, from (5),

$$\left\{\begin{aligned}(p+1)^{-1}\left|\sum_n^{n+p} b_m\right|&\leqq\frac{1}{2\pi}\int_0^{\varepsilon}|T|\,|dH(\vartheta)|+\frac{1}{2\pi}\int_{\varepsilon}^{2\pi}|T|\,|dH(\vartheta)|\\&\leqq\frac{1}{2\pi}\int_0^{\varepsilon}|dH(\vartheta)|+(2\pi)^{-1}(p+1)^{-1}\operatorname{cosec}\frac{1}{2}\varepsilon\int_0^{2\pi}|dH(\vartheta)|.\end{aligned}\right. \tag{6}$$

Since $H(\vartheta)$ is continuous and ε was arbitrary it follows that the left hand side of (6) tends to zero uniformly with respect to n, i.e. $\{b_n\}$ is almost convergent to zero. We now form the convolution

$$L(\vartheta)=\frac{1}{2\pi}\int_0^{2\pi} H(\vartheta+t)\,dH(t);$$

then

$$|b_n|^2 = \frac{1}{2\pi}\int_0^{2\pi} e^{in\vartheta}\, dL(\vartheta). \tag{7}$$

Obviously, since $H(\vartheta)$ is continuous, $L(\vartheta)$ is also continuous. But we have just shown that for a continuous $H(\vartheta)$ the sequence $\{b_n\}$ is almost convergent to zero and consequently, by (7), replacing $H(\vartheta)$ by $L(\vartheta)$, b_n by $|b_n|^2$, it follows that $\{|b_n|^2\}$ is almost convergent to zero. (The fact that, for a continuous $H(\vartheta)$, $\{|b_n|^2\}$ does not, in general, converge to zero is illustrated by the (continuous) Riesz function $F(\vartheta)$ defined below.)

We next prove that $\{|a_n|^2\}$ is almost convergent to $(4\pi)^{-2}\sum \mu_m^2$. By (3) and the definition of $G(\vartheta)$,

$$a_n = \frac{1}{2\pi}\sum_1^\infty \mu_m e^{i\lambda_m n}, \tag{8}$$

$$\frac{1}{2\pi}\sum_1^\infty |\mu_m| = \alpha \text{ say, where } \alpha < \infty. \tag{9}$$

By Lemma 1 each of the terms of the series in (8) is almost periodic in n; by (9) and Lemma 2 it follows that $\{a_n\}$ is almost periodic and so, by Lemmas 3 and 4, $\{|a_n|^2\}$ is almost convergent. It is now enough to show that $\{|a_n|^2\}$ is $(C, 1)$ summable to $(4\pi)^{-2}\sum \mu_m^2$. For integer M let

$$s_M(n) = \frac{1}{2\pi}\sum_1^M \mu_m e^{i\lambda_m n}, \tag{10}$$

$$\frac{1}{2\pi}\sum_{M+1}^\infty |\mu_m| = \varepsilon_M. \tag{11}$$

Then, by (8), (10) and (11),

$$|s_M(n)| - \varepsilon_M \leqq |a_n| \leqq |s_M(n)| + \varepsilon_M.$$

By (9) and (10), since $|s_M(n)| \leqq \sum|\mu_m|$, on squaring these inequalities, summing from 0 to p and dividing by $p+1$, we obtain

$$\left\{\begin{aligned} (p+1)^{-1}\sum_0^p |s_M(n)|^2 - \varepsilon_M(2\alpha+\varepsilon_M) &\leqq (p+1)^{-1}\sum_0^p |a_n|^2 \\ &\leqq (p+1)^{-1}\sum_0^p |s_M(n)|^2 + \varepsilon_M(2\alpha+\varepsilon_M). \end{aligned}\right. \tag{12}$$

Let $(p+1)^{-1}\sum_0^p e^{i(\lambda_l-\lambda_m)n} = \eta(p; l, m)$. Then

$$(p+1)^{-1}\sum_0^p |s_M(n)|^2 = \frac{1}{4\pi^2}\sum_1^M \mu_m^2 + \frac{1}{4\pi^2}\sum_{l=1}^M \sum_{m=1}^{M}{}' \mu_l\mu_m \eta(p; l, m), \tag{13}$$

where the prime indicates that terms for which $l=m$ are excluded. Now if $l \neq m$, we have

$$\eta(p; l, m) = o(1); \tag{14}$$

it follows from (12), (13) and (14) that

$$\frac{1}{4\pi^2}\sum_1^M \mu_m^2 - \varepsilon_M(2\alpha + \varepsilon_M) \leqq \underline{\lim}\,(p+1)^{-1}\sum_0^p |a_n|^2$$

$$\leqq \overline{\lim}\,(p+1)^{-1}\sum_0^p |a_n|^2 \leqq \frac{1}{4\pi^2}\sum_1^M \mu_m^2 + \varepsilon_M(2\alpha + \varepsilon_M),$$

and finally by (9) and (11), on allowing $M \to \infty$, that $\{|a_n|^2\}$ is $(C, 1)$ summable to $(4\pi)^{-2}\sum \mu_m^2$. This completes the proof of part (i) of the theorem.

To prove part (ii), suppose first that $F(\vartheta)$ is continuous. Then $\sum \mu_m^2 = 0$, by part (i) $\{|c_n|^2\}$ is almost convergent to zero. and so, therefore, by the Schwarz inequality

$$\sum_n^{n+p} |s_m| < (p+1)^{\frac{1}{2}}\left(\sum_n^{n+p} |s_m|^2\right)^{\frac{1}{2}} \tag{15}$$

applied to $s_n = c_n$, is $\{|c_n|\}$. Suppose now that $\{|c_n|^2\}$ is summable to zero by some summation method containing almost convergence. By part (i), since $\{|c_n|^2\}$ is almost convergent to $(4\pi)^{-2}\sum \mu_m^2$ it follows that $\sum \mu_m^2 = 0$, $\mu_m = 0$ for all m and $F(\vartheta)$ is continuous. Suppose finally that $\{|c_n|\}$ is summable to zero by such a method. We have already found that $\{|b_n|^2\}$ is almost convergent to zero; by the inequality (15) with $s_n = b_n$, so then is $\{|b_n|\}$. We also found that $\{a_n\}$ is almost periodic; by Lemmas 3 and 4 it follows that $\{|a_n|\}$ is almost convergent to s, say. From the inequalities

$$|a_n| - |b_n| \leqq |c_n| \leqq |a_n| + |b_n|$$

we deduce that $\{|c_n|\}$ is itself almost convergent to s. By (1) the coefficients c_n are bounded, so that $|c_n| < K$, say, and

$$(p+1)^{-1}\sum_n^{n+p} |c_m|^2 \leqq K(p+1)^{-1}\sum_n^{n+p} |c_m|$$

Since s must be zero this inequality implies that $\{|c_n|^2\}$ is almost convergent to zero. An appeal to part (i) shows, as above, that $F(\vartheta)$ is continuous, and this completes the proof.

In conclusion we would point out that if $H(\vartheta)$ is absolutely continuous then $\lim b_n = 0$ and, by Lemma 3, $\{a_n\}$ being almost periodic, $\{c_n\}$, $\{|c_n|\}$ and $\{|c_n|^2\}$ are almost periodic. It is natural to ask whether this is true quite generally of any of the latter sequences. A counterexample is provided by

F. RIESZ's function [3]

$$F(\vartheta) = \lim_{m \to \infty} \int_0^{\vartheta} \prod_{p=1}^{m} (1 + \cos 4^p t)\, dt.$$

For this function, $c_n = 1$ if n is of the form 4^p, but otherwise $c_n \leqq \frac{1}{2}$. Clearly none of the above mentioned sequences is almost periodic.

References

[1] LORENTZ, G. G.: A contribution to the theory of divergent sequences. Acta Math. **80**, 167–190 (1948). – [2] WIENER, N.: The quadratic variation of a function and its Fourier coefficients. Massachusetts J. Math. **3**, 72–94 (1924); see, e.g., ZYGMUND, A.: Trigonometrical Series, § 9.7. – [3] RIESZ, F.: Über die Fourierkoeffizienten einer stetigen Funktion von beschränkter Schwankung. Math. Z. **2**, 312–315 (1918); see, e.g., A. ZYGMUND, l.c., § 5.7.

University College of Swansea

University of New Mexico

(Eingegangen am 18. Juni 1958)

HOFMANN, K. H.
Math. Zeitschr. 71, 36–68 (1959)

Topologische distributive Doppelloops

Von

KARL HEINRICH HOFMANN

Einleitung

Die für die Grundlagen der Geometrie wichtige Theorie der Ternärkörper, also im wesentlichen körperartiger Strukturen ohne Assoziativgesetze für Addition und Multiplikation, ist bislang noch nicht unter einem einheitlichen Gesichtspunkt darstellbar, so daß man zunächst darauf angewiesen ist, spezielle Teilklassen näher zu untersuchen. Wir werden uns im folgenden dem Studium unendlicher topologischer distributiver Doppelloops zuwenden. Eine erste Vorstellung solcher topologisch-algebraischer Strukturen gewinnt man, indem man von den Axiomen, die einen topologischen Körper kennzeichnen, die Assoziativität und Kommutativität sowohl der Addition als auch der Multiplikation wegläßt.

Das Hauptaugenmerk soll dabei derjenigen Teilklasse vorbehalten bleiben, bei deren Vertretern wenigstens die Multiplikation assoziativ ist. Derartige Doppelloops nennt man Neokörper. Bezüglich der von uns betrachteten Topologien werden alle Strukturen lokalkompakt sein. Es ist bekannt, daß topologische Doppelloops entweder zusammenhängend oder total unzusammenhängend sind [5] (1.3). Die total unzusammenhängenden Neokörper werden sich nicht sehr übersichtlich erfassen lassen; immerhin werden wir feststellen können, daß sie hinsichtlich ihrer Topologien und ihrer Multiplikationsgruppe eine gewisse Verwandtschaft zu den nichtarchimedisch bewerteten Körpern aufweisen. Hingegen werden wir die Frage nach den lokalkompakten zusammenhängenden endlichdimensionalen topologischen Neokörpern vollständig beantworten. Diese sind nämlich eins-, zwei- oder vierdimensional und ihre multiplikative Gruppe ist isomorph zu der multiplikativen Gruppe eines der klassischen Pontrjaginschen Schiefkörper. Später zeigen wir darüber hinaus, daß ein lokalkompakter zusammenhängender topologischer Neokörper mit diassoziativer Addition ein Schiefkörper ist. Wesentlich schwächere algebraische Bedingungen werden genügen, um bei etwas schärferen topologischen Voraussetzungen den Körper der reellen bzw. der komplexen Zahlen zu kennzeichnen.

Diese Untersuchungen sind, abgesehen von den Anwendungen in der Geometrie, wo man die Ebenen über Neokörpern als Koordinatenternärkörper geometrisch behandeln kann, vielleicht auch axiomatisch von Interesse, da sie zu den klassischen Schiefkörpern vorstoßen, ohne auf langen Strecken über die Addition irgend etwas vorauszusetzen, es sei denn die Loopeigenschaft.

2061

§ 1. Algebraische Grundbegriffe

(1.1) Unter einer *Doppelloop* versteht man nach PICKERT [*13*], S. 61 eine algebraische Struktur $(K, +, \cdot)$ mit zwei Verknüpfungen $(x, y) \to x + y$ und $(x, y) \to xy = x \cdot y$, die man als *Addition* bzw. *Multiplikation* bezeichnet und die folgenden Bedingungen genügen

a) $(K, +)$ ist eine Loop mit neutralem Element **0**.

b) $(K - \{\mathbf{0}\}, \cdot)$ ist eine Loop mit neutralem Element **1**.

c) $\mathbf{0}x = x\mathbf{0} = \mathbf{0}$ für alle $x \in K$.

Die Loop $(K - \{\mathbf{0}\}, \cdot)$ bezeichnen wir stets mit K'. Eine *rechtsdistributive Doppelloop* ist eine Doppelloop in der

d) $(x + y)z = xz + yz$ für alle $(x, y, z) \in K \times K \times K$

gilt; eine *linksdistributive* ist durch

e) $z(x + y) = zx + zy$ für alle $(x, y, z) \in K \times K \times K$

gekennzeichnet; und schließlich bezeichnen wir eine Doppelloop, in der d) und e) gilt, als *distributiv*. Bei Gültigkeit von d) erzeugen alle Rechtsmultiplikationen $x \to xa$ mit $a \in K'$ eine Automorphismengruppe von $(K, +)$, weshalb wir die Bezeichnung „rechtsdistributives Gesetz" für d) im Gegensatz zu PICKERT ([*14*], S. 90) vorziehen. Bei HUGHES [*6*], S. 507 wird eine distributive Doppelloop ein Divisionsneoring genannt Eine wesentliche Verschärfung der Bedingung b) ist die folgende Aussage

b') $(K - \{\mathbf{0}\}, \cdot)$ ist eine Gruppe mit neutralem Element **1**.

Eine algebraische Struktur $(K, +, \cdot)$ die die Axiome a), b'), c), d) und e) erfüllt nennt man einen *Neokörper* (vgl. PAIGE [*12*], S. 40). Ist K' insbesondere eine abelsche Gruppe, so wird sie als *kommutativer Neokörper* bezeichnet. Es ist zu beachten, daß die Addition dabei von keiner Einschränkung betroffen wird. Eine Loop heißt *potenzassoziativ* (bzw. *monassoziativ*), wenn jedes ihrer Elemente in einer Gruppe (bzw. in einer Halbgruppe liegt); liegen sogar zwei beliebige Elemente stets in einer Gruppe, so nennen wir sie *diassoziativ*. Eine rechtsdistributive Doppelloop hat eine potenzassoziative (bzw. eine monassoziative) Addition genau dann, wenn ein von **0** verschiedenes Element (z. B. das Element **1**) in einer Gruppe (bzw. Halbgruppe liegt).

Nun sei $(H, +, \cdot)$ eine Unterdoppelloop von $(K, +, \cdot)$. Unter dem *Zentrum* der Loop K' versteht man bekanntlich die Menge aller Elemente x von K' für welche

f) $ax = xa$ für alle $a \in K'$,

g) $(ab)x = a(bx)$ für alle $(a, b) \in K' \times K'$,

h) $(ax)b = a(xb)$ für alle $(a, b) \in K' \times K'$,

i) $(xa)b = x(ab)$ für alle $(a, b) \in K' \times K'$

gilt, s. z. B. BRUCK [*1*], S. 255. Ist gerade H' das Zentrum von K', so nennen wir $(H, +, \cdot)$ das *Zentrum* von $(K, +, \cdot)$. Ist indessen H' die Menge der Elemente x, die den Bedingungen g), h) und i) allein genügt, so nennen wir H' den *Kern* von K' und $(H, +, \cdot)$ den *Kern* von $(K, +, \cdot)$. Der Kern von $(K, +, \cdot)$

ist ein Neokörper und das Zentrum von $(K, +, \cdot)$ ein kommutativer Neokörper. Der Kern umfaßt das Zentrum.

(1.2) *Eine distributive Doppelloop* $(K, +, \cdot)$ *besitzt einen Kern* $(C, +, \cdot)$ *und ein Zentrum* $(Z, +, \cdot)$. (Vgl. auch HUGHES [6], S. 508.)

Beweis. Es sei C' der Kern von K' und Z' das Zentrum von K'. Bezeichnen wir $C' \cup \{0\}$ mit C und $Z' \cup \{0\}$ mit Z, so sind C und Z Unterloops von $(K, +)$ bezüglich der Addition. Wir beweisen die Behauptung z.B. für Z; für C folgt sie analog. Z ist genau die Menge aller Elemente, die, bezeichnet man die Rechtsmultiplikationen $x \to xa$ mit R_a und die Linksmultiplikationen $x \to ax$ mit L_a, bei allen Abbildungen $R_a R_b R_{ab}^{-1}$ und $R_b L_a R_{ab}^{-1}$ festbleiben (BRUCK [1], S. 259).

Infolge der Distributivgesetze sind R_a, R_b, L_b und R_{ab} Automorphismen von $(K, +)$. Mit x und y bleibt dann aber auch $z = x + y$ bei jenen Abbildungen fest; bleibt umgekehrt x und z (bzw. y und z) fest, dann auch notwendig y (bzw. x). Damit ist die Behauptung bewiesen. Nun sind also $(C, +, \cdot)$ und $(Z, +, \cdot)$ Unterdoppelloops von $(K, +, \cdot)$. Da C und Z mindestens $\mathbf{0}$ und $\mathbf{1}$ enthalten, sind diese nicht trivial.

(1.3) *Liegt ein von* $\mathbf{0}$ *verschiedenes Element einer distributiven Doppelloop* $(K, +, \cdot)$ *in einer additiven Gruppe, so enthält ihr Zentrum* $(Z, +, \cdot)$ *einen Unterkörper, der einem Primkörper der Charakteristik* p *isomorph ist.*

Beweis. a) Es gebe eine natürliche Zahl n, so daß $n \cdot \mathbf{1} = (\mathbf{1} + \cdots + \mathbf{1}) = \mathbf{0}$ gilt. Die natürlichen Zahlen sind als Operatoren von K in diesem Sinne zulässig, da $(K, +)$ potenzassoziativ ist.

Wir wählen n minimal. Dann ist $n = p$ eine Primzahl; denn aus $n = p \cdot q$, p und q positiv ganz, folgt wegen der Distributivität $\mathbf{0} = n \cdot \mathbf{1} = p \cdot (q \cdot \mathbf{1}) = (p \cdot \mathbf{1})(q \cdot \mathbf{1})$, also z.B. $p \cdot \mathbf{1} = \mathbf{0}$. Die Minimalität von n zieht aber $n = p$ und $q = 1$ nach sich. Die Abbildung $m \to m \cdot \mathbf{1}$ erkennt man rasch als Homomorphismus des Rings N der ganzen Zahlen in $(K, +, \cdot)$. Der Kern dieses Homomorphismus ist das Ideal (p) und das Bild ist isomorph zum Primkörper der Charakteristik p und ist die von $\mathbf{1}$ erzeugte Unterdoppelloop und liegt daher in Z.

b) Es sei $n \cdot \mathbf{1} \neq \mathbf{0}$ für alle ganzen n. Dabei ist $(-n) \cdot \mathbf{1} = n \cdot (-\mathbf{1})$ für ein natürliches n in der von $\mathbf{1}$ erzeugten Gruppe erklärt. Die Abbildung $n \to n \cdot \mathbf{1}$ ist nun ein Isomorphismus von N in $(K, +, \cdot)$. Das Bild liegt in der von $\mathbf{1}$ erzeugten Unterdoppelloop und liegt daher in Z. Sind l, n und m ganze Zahlen, so gilt $(n \cdot \mathbf{1})^{-1}(m \cdot \mathbf{1}) = (n \cdot \mathbf{1})^{-1}(r \cdot \mathbf{1})^{-1}(r \cdot \mathbf{1})(m \cdot \mathbf{1}) = (rn \cdot \mathbf{1})^{-1}(rm \cdot \mathbf{1})$, da ja $(r \cdot \mathbf{1})(m \cdot \mathbf{1}) = rm \cdot \mathbf{1}$ gilt. Einer rationalen Zahl m/n läßt sich also eindeutig das Element $(n \cdot \mathbf{1})^{-1}(m \cdot \mathbf{1}) = f(m/n) \in Z$ zuordnen. f ist bezüglich der Multiplikation ein Isomorphismus des Körpers der Rationalzahlen in Z. Dies gilt auch für die Addition. Es seien nämlich $r_1 = p_1/q$ und $r_2 = p_2/q$ zwei Rationalzahlen mit dem Hauptnenner q. Dann gilt $f(r_1 + r_2) = f\left(\frac{p_1 + p_2}{q}\right) = (q \cdot \mathbf{1})^{-1}((p_1 + p_2) \cdot \mathbf{1}) = (q \cdot \mathbf{1})^{-1}(p_1 \cdot \mathbf{1} + p_2 \cdot \mathbf{1}) = (q \cdot \mathbf{1})^{-1}(p_1 \cdot \mathbf{1}) + (q \cdot \mathbf{1})^{-1}(p_2 \cdot \mathbf{1}) = f(r_1) + f(r_2)$. Infolge der Distributivität wird also der Körper der Rationalzahlen isomorph in Z abgebildet.

(1.4) Liegt ein von **0** verschiedenes Element einer distributiven Doppelloop in einer Halbgruppe, so enthält ihr Zentrum entweder einen Unterkörper isomorph zu einem Primkörper der Charakteristik $p>0$ oder eine distributive Unterstruktur $(H, +, \cdot)$ derart, daß $(H-\{\mathbf{0}\}, \cdot)$ der Gruppe aller positiven rationalen Zahlen und $(H, +)$ der Halbgruppe aller nicht negativen rationalen Zahlen isomorph ist. Der Beweis benötigt gegenüber dem Beweis von (1.2) keine neuen Überlegungen und wird daher weggelassen.

Die Sätze (1.2) bis (1.4) lassen noch Verallgemeinerungen für den Fall zu, daß nur ein Distributivgesetz vorausgesetzt wird. Die Menge H' aller Elemente x in einer Loop K', die der Forderung (1.1) i) genügen, bilden eine Untergruppe, die wir als *Rechtskern* bezeichnen (BRUCK [1], S. 255). Eine Unterdoppelloop $(H, +, \cdot)$ einer Doppelloop $(K, +, \cdot)$, bei der H' den Rechtskern von K' darstellt, nennen wir deren *Rechtskern*; sie ist im allgemeinen kein Neokörper. Es gilt dann entsprechend der (1.2) der Satz

(1.5) *Eine rechtsdistributive Doppelloop* $(K, +, \cdot)$ *besitzt einen Rechtskern* $(C_R, +, \cdot)$.

Beweis. Wie bei (1.2).

(1.6) Liegt ein von **0** verschiedenes Element einer rechtsdistributiven Doppelloop $(K, +, \cdot)$ in einer additiven Gruppe, so enthält ihr Rechtskern C_R einen Unterkörper, der einem Primkörper der Charakteristik p isomorph ist.

Beweis. a) Der Fall, daß eine natürliche Zahl n mit $n \cdot \mathbf{1} = \mathbf{0}$ existiert, erledigt sich wie in (1.3).

b) Für alle ganzen Zahlen sei $n \cdot \mathbf{1} \neq \mathbf{0}$. Der Ring N der ganzen Zahlen wird durch $n \to n \cdot \mathbf{1}$, infolge der Gültigkeit des Rechtsdistributivgesetzes isomorph in C_R abgebildet. Dieses Bild heiße M, dann ist $M' = (M-\{\mathbf{0}\}, \cdot)$ eine reguläre kommutative Halbgruppe in der Gruppe C_R', die darin eine eindeutig bestimmte abelsche Gruppe erzeugt. In dieser setzen wir wiederum $f(m/n) = (n \cdot \mathbf{1})^{-1}(m \cdot \mathbf{1})$ mit $n \neq 0$. Der Rest des Beweises läuft wie in (1.3).

Wir erwähnen noch einen Satz von PICKERT [14], S. 133.

(1.7) Genügt die Addition eines Neokörpers den Kürzungsregeln

a) $-a + (a + b) = b$,

b) $(b + a) - a = b$

so ist die Addition kommutativ.

(1.8) Der Begriff eines Produktes von n Faktoren in einer algebraischen Struktur $(G, \cdot)$ mit einer Verknüpfung $(x, y) \to xy$ ist im allgemeinen nicht eindeutig, sobald G nicht assoziativ ist. Um aber die verschiedenen Möglichkeiten der Multiplikation von n Elementen aus G zu übersehen, greifen wir zu folgender Definition. Zu jeder natürlichen Zahl n geben wir eine endliche Indexmenge J_n und zu jedem Index $i \in J_n$ eine Abbildung f_{ni} der Produktmenge $G \times \cdots \times G$ aus n Faktoren in G an. Wir definieren induktiv

a) J_1 und J_2 bestehen je aus einem Element, und zwar J_1 aus der identischen Abbildung f_1 und J_2 aus der Abbildung f_2 mit $f_2(x, y) = xy$. Es ist also z.B. $f_{1i} = f_1$ mit $i = f_1 \in J_1$.

b) Für $n \geqq 3$ seien die Mengen $J_1, \ldots, J_{n-1}$ definiert und ebenso alle Abbildungen f_{mi} mit $i \in J_m$ und $1 \leqq m \leqq n-1$. Als Elemente der Menge J_n betrachten wir alle Paare $j = (f_{li}, f_{l'i'})$ mit $l + l' = n$, $0 < l < n$, $i \in J_l$, $i' \in J_{l'}$ und setzen $f_{nj}(x_1, \ldots, x_n) = f_2(f_{li}(x_1, \ldots, x_l), f_{l'i'}(x_{n-l+1}, \ldots, x_n))$.

Die Funktionen f_{ni}, $i \in J_n$ geben sozusagen alle möglichen Beklammerungen eines Produkts aus n Faktoren an. G ist genau dann assoziativ, wenn $f_{ni} = f_{ni'}$ für alle n und $(i, i') \in J_n \times J_n$ gilt. $x \in G$ liegt genau dann in einer Gruppe, wenn $f_{ni}(x, \ldots, x) = f_{ni'}(x, \ldots, x)$ gilt für alle n und $(i, i') \in J_n \times J_n$.

Als Randbemerkung sei angegeben, wie man die maximale Elementezahl a_n von J_n leicht berechnet. Es ist $a_2 = a_1 = 1$ und $a_n = \sum_{\substack{l+l'=n \\ 0<l<n}} a_l a_{l'}$, für $n \geqq 3$.

Setzen wir im Körper aller formalen Laurent-Reihen über dem Körper R der reellen Zahlen $p(x) = x + x^2 + a_3 x^3 + a_4 x^4 + \cdots$, so ist $p(x)^2 = x^2 + 2x^3 + \left(\sum_{\substack{l+l'=4 \\ 0<l<4}} a_l a_{l'}\right) x^4 + \cdots$. Es folgt daraus $p(x)^2 - p(x) + x = 0$ oder $p(x) = \frac{1}{2} \cdot \left(1 \pm \sqrt{1 - 4x}\right) = \frac{1}{2} \cdot \left(1 \pm \sum_{l=0}^{\infty} \binom{\frac{1}{2}}{l} (-4x)^l\right)$. Die Lösung mit dem negativen Vorzeichen liefert das gewünschte Resultat, nämlich $a_n = \frac{1}{n}\binom{2(n-1)}{n-1}$ für $n \geqq 3$.

(1.9) In einer früheren Arbeit des Verfassers ist der Begriff des Fastvektorraums eingeführt worden [4]. Wir geben diese Definition in etwas allgemeinerer Form hier wieder: $(G, +)$ sei eine additiv geschriebene Loop. $(K, +, \cdot)$ sei eine Doppelloop. (Man unterscheide die Additionen in G und in K!) Wir nennen G einen *K-Fastverktorraum*, wenn folgende Bedingungen erfüllt sind

a) Die Elemente von K sind Endomorphismen von G in sich. Ein solcher Endomorphismus wird mit $x \to r \cdot x$ bezeichnet.

b) Aus $(r, s) \in K \times K$ und $x \in G$ folgt $(r + s) \cdot x = r \cdot x + s \cdot x$ und $(rs) \cdot x = r \cdot (s \cdot x)$.

c) $1 \cdot x = x$.

Aus b) folgt $\mathbf{0} \cdot x = \mathbf{0}$ und die Assoziativität der Multiplikation in K. Jeder Endomorphismus $r \in K$, $r \neq \mathbf{0}$ hat eine Umkehrabbildung, nämlich r^{-1} und ist daher ein Automorphismus. Sind p, q, s Elemente aus K, so gilt wegen b) und der Endomorphieeigenschaft $p(r + s) \cdot x = p \cdot (r \cdot x + s \cdot x) = p \cdot (r \cdot x) + p \cdot (s \cdot x) = pr \cdot x + ps \cdot x = (pr + ps) \cdot x$. Ebenso zeigt man $(r + s)p \cdot x = (rp + sp) \cdot x$. Daraus folgen die beiden Distributivgesetze für K, denn alle von $\mathbf{0}$ verschiedenen Elemente sind Automorphismen von G. Somit ist K ein Neokörper.

(1.10) *Ist $(K, +, \cdot)$ eine beliebige distributive Doppelloop, so ist $(K, +)$ ein Z- und ein C-Fastvektorraum, wenn Z das Zentrum und C den Kern von K bezeichnen.*

(1.11) *$(G, +)$ ist genau dann monassoziativ bzw. potenzassoziativ, wenn K eine monassoziative bzw. potenzassoziative Addition besitzt. Ist $(G, +)$ assoziativ und kommutativ, dann auch $(K, +)$, und G ist K-Fastvektorraum.*

Beweis. Die Abbildung $r \to r \cdot x$ ist ein Isomorphismus von $(K, +)$ in $(G, +)$, wenn $x \neq \mathbf{0}$ ist. Ist also z.B. $(G, +)$ eine abelsche Gruppe, dann auch $(K, +)$, und $(K, +, \cdot)$ ist ein Schiefkörper.

(1.12) Unter dem *Erzeugnis* einer Menge M von Elementen in einem K-Fastvektorraum G verstehen wir den kleinsten K-Unterfastvektorraum von G, der M enthält. Das Erzeugnis eines Elementes x ist $K \cdot x$.

(1.13) *$(G, +)$ sei ein $(K, +, \cdot)$-Fastvektorraum. Dann sind das Zentrum $Z(G)$, der Kern $C(G)$ und der Rechtskern $C_R(G)$ von G jeweils K-Unterfastvektorräume, $Z(G)$ nach* (1.11) *dann sogar ein Untervektorraum.*

Beweis. Wir beweisen die Behauptung z.B. für $C_R(G)$ und zeigen, daß $x \in C_R(G)$ stets $K \cdot x \subseteq C_R(G)$ nach sich zieht. Die Voraussetzung besagt, daß für alle $(y, z) \in G \times G$ immer $(x + y) + z = x + (y + z)$ gilt. Für ein beliebiges $r \neq \mathbf{0}$ aus K sei x' so bestimmt, daß $(r \cdot x + y) + z = x' + (y + z)$ gilt. Diese Gleichung multiplizieren wir mit r^{-1} durch:

$$r^{-1} \cdot \big((r \cdot x + y) + z\big) = (x + r^{-1} \cdot y) + r^{-1} z$$
$$= r^{-1} \cdot \big(x' + (y + z)\big) = r^{-1} \cdot x' + (r^{-1} \cdot y + r^{-1} \cdot z).$$

Daraus folgt aber nach Voraussetzung $r^{-1} \cdot x' = x$ oder $x' = r \cdot x$, was zu zeigen war.

(1.14) *Hat ein K-Fastvektorraum ein nichttriviales Zentrum $Z(G)$, dann ist K ein Schiefkörper, und $Z(G)$ ist ein Vektorraum über K.*

Beweis. (1.11) und (1.13).

(1.15) *Hat ein K-Fastvektorraum einen nichttrivialen Rechtskern $C_R(G)$, dann ist K ein Schiefkörper.*

Beweis. Nach (1.11) ist $(K, +)$ assoziativ. Dann ist $(K, +)$ aber sogar kommutativ: Siehe z.B. PICKERT [*13*], S. 91.

(1.16) *Jeder Normalteiler eines K-Fastvektorraums liegt in einem normalen K-Unterfastvektorraum.*

Der Beweis folgt dem Schema von (1.12) unter Beachtung der Definition der Normalität, wie sie z.B. in [*1*], S. 252 angegeben ist und wird hier übergangen.

(1.17) *Ist ein K-Fastvektorraum G eine kommutative Moufangloop, so hat K die Charakteristik 3 oder K ist ein Schiefkörper und er selbst sogar ein Vektorraum über K.*

Beweis. $(K, +)$ ist potenzassoziativ (1.12) und K enthält einen Primkörper der Charakteristik p (1.3). Ist $p \neq 3$, so ist $x \to 3 \cdot x$ ein Automorphismus von G, andrerseits aber ein Endomorphismus in sein Zentrum: Siehe z.B. BRUCK [*1*], S. 302. Also ist G assoziativ und kommutativ und K ist ein Schiefkörper (1.11).

Als Korollar folgt hieraus ein Teil eines Satzes von PICKERT [*14*], S. 135, 5:

(1.18) *Ein Neokörper K mit Moufangscher Addition und einer Charakteristik ungleich 3 ist ein Schiefkörper.*

Beweis. Die Addition ist kommutativ nach (1.8). $(K, +)$ ist K-Fastvektorraum. Die Behauptung folgt dann aus (1.17).

(1.19) *Es sei $(K, +, \cdot)$ eine distributive Doppelloop mit einem Element $x \neq 0$, das in einer Untergruppe von $(K, +)$ liegt. R_0 sei der von 1 gemäß (1.6) erzeugte Primkörper von $(K, +, \cdot)$. Es sei x nicht in R_0 enthalten und liege in einer Unterhalbgruppe von $(K', \cdot)$. Ferner sei vorausgesetzt, daß nicht nur x, sondern auch die Menge aller seiner nichtnegativen Potenzen in einer und derselben Untergruppe von $(K, +)$ liege. Dann enthält $(K, +, \cdot)$ einen Integritätsbereich $(H, +, \cdot)$, der x enthält und entweder einer algebraischen Erweiterung von R_0 oder dem Polynomring über R_0 in einer Erzeugenden isomorph ist. Erzeugt also x insbesondere multiplikativ eine endliche Gruppe, so liegt x in einem Unterkörper von K.*

Beweis. Es sei $(H_0, +)$ eine maximale abelsche Untergruppe von $(K, +)$, die alle nicht negativen Potenzen von x enthält. Dann gilt $R_0 H_0 = H_0$, denn ist r ein Element aus R_0, so ist rH_0 eine abelsche Gruppe in $(K, +)$, die in H_0 enthalten ist; jedenfalls ist dies dann richtig, wenn mit einer natürlichen Zahl n die Beziehung $r = n \cdot 1$ gilt. Ist $m \cdot 1 \neq 0$, so ist $(m \cdot 1)^{-1} \cdot H_0$ eine abelsche Gruppe, die H_0 umfaßt, also wegen der geforderten Maximalität von H_0 gleich H_0 ist. Somit liegt aber auch $\big((m^{-1}n) \cdot 1\big)\, H_0 = (m \cdot 1)^{-1} \cdot (n \cdot 1)\, H_0 = (m \cdot 1)^{-1} H_0 = \text{in } H_0$. Demnach ist H_0 sogar ein R_0-Vektorraum. Die Abbildung, die einem Element $\sum_{n=0}^{N} r_n X^n$ des Polynomrings $R_0[X]$ das Element $r_0 \cdot 1 + r_1 \cdot x + \cdots + r_N \cdot x^N$ von H_0 zuordnet, ist ein Homomorphismus. Er bildet $R_0[X]$ auf einen Unterring $(H, +, \cdot)$ von $(K, +, \cdot)$ ab, der nullteilerfrei und daher ein Integritätsbereich ist. Der Kern des Homomorphismus ist ein Hauptideal $(h(X))$ mit einem erzeugenden Primpolynom $h(X)$; andere Polynome kommen wegen der Nullteilerfreiheit des Bildbereichs nicht in Frage. Ist $h(X)$ das Nullpolynom, so ist H isomorph zu $R_0[X]$; im anderen Fall ist wegen $x \notin H_0$ der Grad von $h(X)$ mindestens zwei und H ist eine endliche algebraische Erweiterung von R_0.

Ist insbesondere $(K, +)$ alternativ, und gibt es ein nicht in R_0 liegendes Element x mit $x^2 = -1$, so enthält $(K, +, \cdot)$ einen Körper isomorph zu $R_0(i)$, der x umfaßt.

Liegt x sogar in einer Gruppe in $(K', \cdot)$, und liegen alle Potenzen von x in einer Untergruppe von $(K, +)$, bezeichnet ferner $R_0[X]'$ den Ring aller Polynome $\sum_{n=-N}^{n=N} r_n X^n$, so läßt sich in $(K, +, \cdot)$ sogar ein homomorphes Bild von $R_0[X]'$ finden, das allerdings nur im Fall der Isomorphie größer ist als das bisher konstruierte.

§ 2. Hilfssätze über topologische Loops

(2.1) Eine Loop $(G, \cdot)$ mit der Verknüpfung $(x, y) \to xy$ heißt *topologisch*, wenn auf G eine Topologie so erklärt ist, daß die Abbildungen $(x, y) \to xy$, $(x, y) \to x\backslash y$ und $(x, y) \to x/y$ mit $x(x\backslash y) = y$ und $(x/y)\, y = x$, die den Produkt-

raum $G \times G$ in G abbilden, stetig sind. Wir betrachten im folgenden fast ausschließlich lokalkompakte Loops. Zuerst führen wir einige Hilfssätze auf:

(2.2) $(G, \cdot)$ sei eine total unzusammenhängende lokalkompakte topologische Loop. Wir definieren induktiv $g_1(x) = x$, $g_n(x_1, \ldots, x_n) = g_{n-1}(x_1, \ldots, x_{n-1})\, x$. Dann hat der Umgebungsfilter $\mathfrak{U}(e)$ des neutralen Elements e eine Basis aus den Umgebungen $W_V = \cup \{g_n(V, \ldots, V);\ n = 1, 2, \ldots\}$, $V \in \mathfrak{U}(e)$. Es gilt dann natürlich $W_V \cdot V = W_V$. Den Beweis findet man in einer anderen Arbeit des Verfassers ([4], (1.4)).

Ist $(G, \cdot)$ potenzassoziativ, so liegen die sämtlichen Potenzen aller Elemente von V in W_V.

In der entgegengesetzten Richtung zielen die folgenden Hilfssätze.

Die Bezeichnungen seien im folgenden dieselben wie in (1.8).

(2.3) $(G, \cdot)$ sei eine topologische Loop, U eine beliebige offene Umgebung des neutralen Elementes mit kompakter Hülle $\overline{U}$. Dann gilt für alle natürlichen Zahlen n und alle Indizes i aus J_n

a) $\overline{f_{ni}(U, \ldots, U)} = f_{ni}(\overline{U}, \ldots, \overline{U})$,

b) Ist $W_l = \overline{U}$ für $m \leqq n-1$ der Zahlen $l = 1, \ldots, n$ und $W_l = U$ sonst, dann gilt $f_{ni}(W_1, \ldots, W_n) = f_{ni}(U, \ldots, U)$.

Beweis. a) Falls die Behauptung für $n = 2$ als richtig nachgewiesen ist, folgt sie mittels vollständiger Induktion für beliebige n. In einer topologischen Loop ist aber das Produkt AK aus einem abgeschlossenen Komplex A und einem kompakten Komplex K abgeschlossen (HOFMANN [3], (3.9)). Also gilt $f_2(\overline{U}, \overline{U}) \supseteq \overline{f_2(U, U)}$. Die umkehrte Enthaltenseinsbeziehung ist auf Grund der Stetigkeit von f_2 klar.

b) Es sei $W_k = U$ mit $1 < k < n$. Dann ist $f_{ni}(W_1, \ldots, W_n)$

$$\subseteq \cup \{f_{ni}(x_1, \ldots, x_{k-1}, W_k, x_{k+1}, \ldots, x_n);\ (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n) \in \overline{U} \times \ldots \times \overline{U}\}$$

offen und in $\overline{f_{ni}(U, \ldots, U)}$ enthalten, umfaßt andrerseits aber $f_{ni}(U, \ldots, U)$ und ist daher der offene Kern der Hülle. Dieser ist aber $f_{ni}(U, \ldots, U)$. Eine entsprechende Modifikation der Formulierung liefert die Behauptung auch für $k = 1$ und $k = n$.

(2.4) Es gelten die Voraussetzungen von (2.3). Außerdem habe U einen nichtleeren Rand.

Definieren wir wieder Funktionen g_m durch $g_1(x) = x$ und $g_m(x_1, \ldots, x_m) = g_{m-1}(x_1, \ldots, x_{m-1})\, x_m$ und haben die Mengen $g_n(U, \ldots, U)$ einen nichtleeren Rand für $n = 1, 2, \ldots$, dann gibt es ein x aus $\overline{U}$ derart, daß $g_m(x, \ldots, x) \notin \overline{g_{m-1}(U, \ldots, U)}$ gilt. Da $g_m(U, \ldots, U) - \overline{g_{m-1}(U, \ldots, U)}$ offen ist, hat mit x auch noch jedes y aus einer Umgebung X von x die Eigenschaft $g_m(y, \ldots, y) \notin \overline{g_{m-1}(U, \ldots, U)}$.

Beweis. a) m sei eine feste natürliche Zahl. Zu jedem $i \in J_m$ gibt es eine offene Umgebung V_i des neutralen Elements e, so daß für alle y aus $\overline{U}$ stets

$f_{mi}(y, \ldots, y) \backslash f_{mi}(V_i y, \ldots, V_i y) \subseteq U$ gilt, wobei mit $a \backslash b$ die Lösung von $ax = b$ bezeichnet wurde. Andernfalls gäbe es gegen e konvergente Netze $(v_s^{(l)})_{s \in J}$ mit $l = 1, \ldots, m$ und ein Netz $(y_s)_{s \in J}$ auf $\overline{U}$, so daß immer

$$f_{mi}(y_s, \ldots, y_s) \backslash f_{mi}(v_s^{(1)} y_s, \ldots, v_s^{(m)} y_s) \notin U$$

gilt. Wegen der Kompaktheit von $\overline{U}$ können wir ohne die Allgemeinheit einzuschränken annehmen, daß $(y_s)_{s \in J}$ auf $\overline{U}$ gegen $y \in \overline{U}$ konvergiert; andernfalls könnten wir eine geeignete konvergente Teilfolge finden. Da f_{mi} stetig ist, gilt aber $\lim_{s \in J} f_{mi}(y_s, \ldots, y_s) \backslash f_{mi}(v_s^{(1)} y_s, \ldots, v_s^{(m)} y_s) = e$, während dieser Grenzwert doch im abgeschlossenen Komplement von U liegen müßte. Wir setzen $V = \bigcap_{i \in J} V_i$.

b) Der Kompaktheit von $\overline{U}$ wegen gibt es Elemente $y_1, \ldots, y_n$ in $\overline{U}$, so daß $\overline{U} \subseteq \bigcup_{l=1}^{n} V y_l$ gilt. Aus der Menge der Funktionen f_{ri} sondern wir die uns schon bekannte Klasse von Funktionen g_k aus ($k = 1, \ldots$), indem wir $g_1(x) = x$ und $g_k(x_1, \ldots, x_k) = g_{k-1}(x_1, \ldots, x_{k-1})\, x_k$ setzen. Nun sei $p = mn$. Ist z ein Element des Randes von $g_p(U, \ldots, U)$, so gibt es p Elemente $z_1, \ldots, z_p$ in $\overline{U}$, die wegen (2.2) b) auf dem Rand von $\overline{U}$ liegen, so daß $z = g_p(z_1, \ldots, z_p)$ gilt. Wenigstens eine der Mengen $V y_l$ muß mindestens m der p Elemente $z_1, \ldots, z_p$ enthalten; eine solche Menge sei $V y_1$. In dieser Menge liegen die Elemente $z_1' = z_{i_1}, \ldots, z_m' = z_{i_m}$, wo $1 \leqq i_1 < i_2 < \cdots < i_m \leqq p$. Wir setzen $z_q'' = z_q$, wenn $q \in \{i_1, \ldots, i_m\}$ und $z_q'' = e$ sonst. Dann liegt das Element $z' = g_m(z_1', \ldots, z_m') = g_p(z_1'', \ldots, z_p'')$ auf dem Rand von $g_m(U, \ldots, U)$; denn andernfalls läge nämlich $g_p(z_1'', \ldots, z_p'')$ im Innern von $g_p(A_1, \ldots, A_p)$ mit $A_q = \overline{U}$, wenn $q \in \{i_1, \ldots, i_m\}$ und $A_q = \{e\}$ sonst; dann aber läge $z = g_p(z_1, \ldots, z_p)$ im Innern von $g_p(\overline{U}, \ldots, \overline{U})$ im Widerspruch zur Wahl von z.

c) Für alle l mit $1 \leqq l \leqq m$ gibt es Elemente $v_l \in V$, so daß $z_l' = v_l y_1$. Wir setzen $z'' = g_m(v_1 y_1, \ldots, v_m y_1) = g_m(y_1, \ldots, y_1)\, u$ mit $u \in U$; dies ist nach a) möglich. Läge nun $g_m(y_1, \ldots, y_1)$ in $g_{m-1}(\overline{U}, \ldots, \overline{U})$, so läge $z'' = g_m(y_1, \ldots, y_1)\, u$ nach (2.2) b) in $g_{m-1}(\overline{U}, \ldots, \overline{U}) \cdot U = g_m(U, \ldots, U)$, also nicht, wie es sein muß, auf dem Rand von $g_m(U, \ldots, U)$.

Setzen wir nun noch $x = y_1$, so ist damit die Behauptung bewiesen.

Aus diesem Hilfssatz folgt direkt das Korollar:

(2.5) $(G, \cdot)$ sei eine lokalkompakte monassoziative topologische Loop mit einer relativ kompakten Umgebung U des neutralen Elements mit nichtleerem Rand. Zu jeder natürlichen Zahl m gibt es ein Element x_m derart, daß x_m^m nicht in $(\ldots(UU)\ldots U)\, U = g_{m-1}(U, \ldots, U)$ liegt.

Dieses Ergebnis spezialisieren wir, indem wir $(G, \cdot)$ als zusammenhängend voraussetzen. Dann hat e eine Basis $\mathfrak{U}(e)$ aus offenen, relativ kompakten Umgebungen mit nichtleerem Rand; ferner ist $G = \bigcup_{m=2}^{\infty} g_{m-1}(U, \ldots, U)$ für jede Umgebung $U \in \mathfrak{U}(e)$; denn erstens ist die Menge $H = \bigcup_{m=2}^{\infty} g_{m-1}(U, \ldots, U)$ offen;

andrerseits ist aber die abgeschlossene Hülle $\overline{H}$ in HU enthalten; ist nämlich $x \in \overline{H}$, so gibt es ein $y \in H \cap x/U$ woraus $x \in yU \subseteq HU$ folgt. Nun ist aber $HU = H$; somit ist $H = \overline{H}$ abgeschlossen. Da H nichtleer und G zusammenhängend ist, folgt $G = H$. (Vgl. auch [3], (5.6).) Nun halten wir eine kompakte Umgebung $U = \overline{U}$ fest. Zu jeder relativ kompakten, offenen Umgebung $V \in \mathfrak{U}(e)$ gibt es eine natürliche Zahl $n > 1$ derart, daß $U \subseteq \bigcup_{m=2}^{n} g_{m-1}(V, \ldots, V) = g_{n-1}(V, \ldots, V) = H_n$. Die abgeschlossenen Komplemente $G - H_n$ bilden nämlich eine monoton fallende Folge abgeschlossener Mengen mit leerem Durchschnitt. Es gibt also sicher ein kleinstes n_0 mit $U \cap (G - H_{n_0}) = \emptyset$.

Falls G nichtkompakt ist, dann hat keine der relativ kompakten offenen Mengen $g_n(V, \ldots, V)$ einen leeren Rand.

In $\overline{V}$ wählen wir x_v so, daß eine Potenz von x_v außerhalb U liegt. Dies ist nach (2.4) und dem eben Bewiesenen möglich, und zwar sei m_v die kleinste natürliche Zahl, für welche $x_v^{m_v} \notin U$ gilt. K sei die abgeschlossene Hülle in U der Menge aller x_v^m mit $1 \leqq m \leqq m_v$. Unter diesen Voraussetzungen gibt es einen möglicherweise nicht stetigen Homomorphismus f der additiven Halbgruppe R_+ aller nichtnegativen reellen Zahlen in G derart, daß das Intervall $[0, 1]$ homöomorph in K abgebildet wird und $f(1)$ auf dem Rand von U liegt. Man nennt auch $f([0, 1])$ einen Strahl ([4], (5.2)). $f(R_+)$ heißt Einparameterhalbgruppe.

Ist G sogar potenzassoziativ, so enthält G eine Einparametergruppe, nämlich die Fortsetzung von $f(R_+)$.

Wir haben folgenden Satz bewiesen

(2.6) *Eine zusammenhängende, lokalkompakte nicht kompakte topologische Loop G, die monassoziativ (bzw. potenzassoziativ) ist enthält eine (vielleicht nicht stetige) Einparameterhalbgruppe (bzw. eine Einparametergruppe).*

Statt die Nichtkompaktheit vorauszusetzen, hätte es genügt, z. B. vorauszusetzen, daß G einen nichtkompakten Überlagerungsraum besitzt.

Hat G eine transitive Automorphismengruppe, so liegt in G jeder Punkt auf einer Einparameterhalbgruppe (bzw. auf einer Einparametergruppe).

§ 3. Topologische Doppelloops und ihre Multiplikationsloop bei potenzassoziativer Multiplikation

Unter einer topologischen Doppelloop $(K, +, \cdot)$ verstehen wir die topologische algebraische Struktur, die entsteht, wenn auf der Menge K eine nichtdiskrete Topologie definiert ist, bezüglich der die Additionsloop $(K, +)$ und die Multiplikationsloop $(K', \cdot)$ topologische Loops bilden und bezüglich der die Abbildung $(x, y) \to xy$ auch in allen Punkten (a, b) mit $a = \mathbf{0}$ oder $b = \mathbf{0}$ stetig ist. Zunächst notieren wir einige bekannte Tatsachen. Eine topologische Doppelloop ist entweder zusammenhängend oder nirgends zusammenhängend ([3], (1.3)). Ist sie lokalkompakt und zusammenhängend, so ist sie lokalzusammenhängend (HOFMANN [3], (1.6)). Ist K lokaleinfach zusammenhängend, so ist K einfach zusammenhängend (im Sinne von C. CHEVALLEY [2], S. 44, [5], (1.12)). Eine lokalkompakte Doppelloop K ist nicht kompakt und

besitzt eine homogene Einpunktkompaktifizierung $K \cup \{a\} = \overline{K}$, die Abbildungen $x \to x\backslash \mathbf{1}$ bzw. $x \to \mathbf{1}/x$ haben stetige und umkehrbar stetige Fortsetzungen f bzw. g von $\overline{K}$ auf sich mit $f(\mathbf{0}) = g(\mathbf{0}) = a$ und $f(a) = g(a) = \mathbf{0}$ ([5], (1.5)). Eine lokalkompakte topologische Doppelloop genügt dem ersten Abzählbarkeitsaxiom. Der Beweis kann wörtlich von PONTRJAGIN [15], S. 190 übernommen werden.

Im folgenden sei $(K, +, \cdot)$ eine lokalkompakte Doppelloop. $(K', \cdot)$ sei potenzassoziativ, d. h. jedes Element $x \in K'$ erzeuge eine zyklische Gruppe $\langle x \rangle$. Diese ist entweder unendlich diskret, oder $\overline{\langle x \rangle}$ ist kompakt (vgl. z. B. WEIL [17], S. 96). Die Menge $\{x; \langle x \rangle$ unendlich diskret$\}$ heiße A, die Menge $\{x; \overline{\langle x \rangle}$ kompakt$\}$ heiße B. Ist $x \in A$, so hat $\langle x \rangle$ in K' keine Häufungspunkte, muß aber in $\overline{K}$ solche besitzen. Wegen $f(\langle x \rangle) = \langle x \rangle$ sind $\mathbf{0}$ und a Häufungspunkte von $\langle x \rangle$ in $\overline{K}$. Wir behaupten sogar $\lim\limits_{n \to \infty} x^n = \mathbf{0}$ oder $\lim\limits_{n \to \infty} x^n = a$. Falls diese Behauptung richtig ist, sind $\lim\limits_{n \to \infty} x^n = \mathbf{0}$ und $\lim\limits_{n \to \infty} x^{-n} = a$ gleichwertig. Die Menge $\{x;\ x = \mathbf{0}$ oder $\lim\limits_{n \to \infty} x^n = \mathbf{0}\}$ nennen wir dann A_0, und die Menge $\{x;\ x = a$ oder $\lim\limits_{n \to \infty} x^n = \mathbf{0}\} = f(A_0)$ heiße A_a, so daß also dann $A \cup \{0\} \cup \{a\} = A_0 \cup A_a$ gilt. Zum Beweis unserer Behauptung nehmen wir an, es gäbe Folgen $(n_i)_{i=1,\ldots}$ und $(m_i)_{i=1,\ldots}$ natürlicher Zahlen mit $\lim\limits_{i \to \infty} x^{n_i} = a$ und $\lim\limits_{i \to \infty} x^{m_i} = \mathbf{0}$. Die Folge $(x^{n_i - m_i})_{i=1,\ldots}$ hat in $\overline{K}$ einen Häufungspunkt, der etwa gleich $\mathbf{0}$ sei (wäre nur a Häufungspunkt, so könnte man durch Anwendung der Funktion f die gewünschte Aussage erzwingen). Da wir zu einer konvergenten Teilfolge übergehen können, können wir schließlich sogar $\lim\limits_{i \to \infty} x^{n_i - m_i} = \mathbf{0}$ annehmen. Damit ist aber $\mathbf{0} = \mathbf{0} \cdot \mathbf{0} = \left(\lim\limits_{i \to \infty} x^{n_i - m_i}\right) \cdot \left(\lim\limits_{i \to \infty} x^{m_i}\right) = \lim\limits_{i \to \infty} (x^{n_i - m_i} \cdot x^{m_i}) = \lim\limits_{i \to \infty} x^{n_i} = a$. Dieser Widerspruch beweist die Behauptung. Ist nun $x \in B$, so ist auch $x^{-1} \in B$; da $\overline{\langle x \rangle}$ kompakt ist, ist $\overline{K} - \overline{\langle x \rangle}$ eine Umgebung von a. Daher ist a und wegen $f(\overline{\langle x \rangle}) = \overline{\langle x \rangle}$ auch $\mathbf{0}$ kein Adhärenzpunkt von x. Der Raum $\overline{K}$ ist also zerlegt in der Form $A_0 \cup B \cup A_a$ und diese Zerlegung ist invariant gegen Anwendung von f. Es sei nun W eine offene, relativ kompakte Umgebung von $\mathbf{0}$, deren Hülle $\mathbf{1}$ nicht enthält. Wir setzen $V = (W \cup \{\mathbf{1}\})(W \cup \{\mathbf{1}\})$. Einem Gedanken von PONTRJAGIN folgend beweisen wir, daß aus der Existenz einer natürlichen Zahl n mit $g_{n+1}(V, x, \ldots, x) \subseteq W$, wobei g_m die schon in (2.3) verwendeten Funktion bezeichnen, stets $x \in A_0$ folgt. Jene Bedingung ist in der Tat hinreichend: Denn ist n eine natürliche Zahl, für welche $g_{n+1}(V, x, \ldots, x) \subseteq W$ gilt, so ist für alle natürlichen m auch die Bedingung $g_{nm+1}(V, x, \ldots, x) \subseteq W$ richtig. Denn ist sie etwa für m_0 richtig, so folgt $g_{n(m_0+1)+1}(V, x, \ldots, x) = g_{n+1}(g_{nm_0+1}(V, x, \ldots, x), x, \ldots, x) \subseteq g_{n+1}(W, x, \ldots, x) \subseteq W$, also ihre Richtigkeit für $m_0 + 1$. Da V das Element $\mathbf{1}$ enthält, folgt daraus $x^{nm} \in W$ für alle natürlichen m. Läge x in B, dann auch x^n und W enthielte $\{x^{nm}, m = 1, 2, \ldots\}$. Da x^n eine relativ kompakte Gruppe erzeugt, ist schon die Menge aller positiven Potenzen von x^n in $\overline{\langle x^n \rangle}$ dicht. Es folgt $\overline{\langle x^n \rangle} \subseteq \overline{W}$ und damit insbesondere $\mathbf{1} \in \overline{W}$ im Widerspruch zur Wahl von W. Demnach liegt x in $A_0 \cup A_a$; es gilt aber weder $x = a$ noch $\lim\limits_{m \to \infty} x^{nm} = a$, also liegt x wirklich in A_0.

Nun sei x ein Element, für welches $g_{n+1}(V, x, \ldots, x) \subseteq W$ gilt. Zu jedem $y \in V$ gibt es Umgebungen U_y von y und U_y' von x derart, daß auch $g_{n+1}(U_y, U_y', \ldots, U_y') \subseteq W$ gilt. V wird von endlich vielen Umgebungen $U_{y_1}, \ldots, U_{y_l}$ überdeckt. Setzen wir $U = \bigcap_{m=1}^{l} U_{y_m}'$, so gilt auch $g_{n+1}(V, U, \ldots, U) \subseteq W$. Daraus folgt dann $U \subseteq A_0$. Setzen wir insbesondere $x = \mathbf{0}$, so folgt, daß A_0 eine Umgebung von $\mathbf{0}$ und damit dann auch A_a eine Umgebung von a ist; $\overline{B}$ ist somit aber kompakt.

Wir betrachten jetzt ein Element x aus $A_0 \cap K'$. Wegen $\lim_{n\to\infty} x^n = \mathbf{0}$ gibt es ein m, so daß x^m im Innern von A_0 liegt. Wegen der Stetigkeit von $y \to y^m$ gibt es eine Umgebung X von x, so daß für alle $y \in X$ auch y^m im Innern von A_0 liegt. Dann ist aber $X \subseteq A_0$ und A_0 ist offen. Damit ist auch A_a offen. Wir haben folgenden Hilfssatz bewiesen:

(3.1) *$(K, +, \cdot)$ sei eine lokalkompakte topologische Doppelloop mit potenzassoziativer Multiplikation. Dann zerfällt K in drei paarweise disjunkte Mengen A_0, A_a und B, von denen A_0 und A_a offen sind, während B kompakt ist. Es gilt* $\mathbf{0} \in A_a$, $\mathbf{1} \in B$, $f(A_0) = g(A_0) = A_a$ und $f(B) = g(B) = B$.

Aus $x \in A_0$ folgt $\lim_{n\to\infty} x^n = \mathbf{0}$; aus $x \in A_a$ folgt $\lim_{n\to\infty} x^n = a$, für alle $x \in B$ ist $\overline{\langle x \rangle}$ eine kompakte Gruppe.

Hat K ein Zentrum Z und gilt $Z \cap A_0 \neq \emptyset$, so gelten die Beziehungen $A(Z \cap A) = A$ und $B(Z \cap B) = B$, sowie $A(Z \cap B) \subseteq A$ und $B(Z \cap A) \subseteq A$; denn ist etwa $x \in A$ und $y \in Z \cap B$, so ist $\langle xy \rangle = \{x^n y^\nu;\ n = 0, \pm 1, \ldots\}$ nicht relativ kompakt, also liegt xy in A. Die Menge $Z_2 = Z \cap B$ ist eine kompakte abelsche Gruppe. Mit Hilfe des Zornschen Lemmas findet man in der abelschen Gruppe $(Z', \cdot)$ einen torsionsfreien Komplementärfaktor Z_1 mit $Z_1 Z_2 = Z'$ und $Z_1 \cap Z_2 = \{\mathbf{1}\}$: Eine maximale Untergruppe Z_1 von Z' mit der Eigenschaft $Z_1 \cap Z_2 = \{\mathbf{1}\}$ leistet das Verlangte; wäre nämlich $z \in A \cap Z'$ nicht in $Z_1 Z_2$, so läge das Produkt $Z_1 \cdot \langle z \rangle$ in $A \cup \{\mathbf{1}\}$ und hätte mit Z_2 den Durchschnitt $\{\mathbf{1}\}$. Wir zeigen, daß Z_1 sogar abgeschlossen gewählt werden kann. Zunächst sei Z zusammenhängend und dann auch lokalzusammenhängend. Dann wählen wir Z_1' maximal zusammenhängend mit $Z_1' \cap Z_2 = \{\mathbf{1}\}$. Die Zusammenhangskomponente von $\mathbf{1}$ in Z' wird dann von $Z_1' Z_2$ ganz überdeckt. Denn ist z ein in dieser Komponente, nicht aber in $Z_1' Z_2$ gelegenes Element, so geht durch z eine nicht relativ kompakte Einparametergruppe E. $Z_1' E$ liegt in $A \cup \{\mathbf{1}\}$ und ist zusammenhängend. Wegen der Maximalität von Z_1' ist $Z_1' E = Z_1$, woraus $z \in Z_1' = Z_1' Z_2$ folgen müßte. $\overline{Z_1'}$ ist eine zusammenhängende lokalkompakte abelsche Gruppe, in der eine Gruppe ohne kompakte Untergruppen dicht liegt. Sie ist demnach isomorph dem euklidischen Vektorraum R^n für geeignet n.

$\overline{Z_1'} \cap Z_2$ ist eine kompakte Untergruppe von $\overline{Z_1'}$ und daher gleich $\{\mathbf{1}\}$. Wiederum folgt aus der Maximalität $Z_1' = \overline{Z_1'}$.

Da Z' das erste Abzählbarkeitsaxiom erfüllt, ist die Zusammenhangskomponente Z_0' von $\mathbf{1}$ in Z' isomorph $R^m \times T^m$, wo T^m eine m-dimensionale Torusgruppe bezeichnet, $0 \leq m \leq \infty$. Da Z lokalzusammenhängend ist, ist

diese offen in Z'. Nun ist Z'_0 eine Untergruppe mit Division; als solche ist sie algebraisch ein direkter Faktor von Z'. Da sie offen ist, ist der Komplementärfaktor Y diskret und damit abgeschlossen, d.h. Z'_0 ist auch topologisch ein direkter Faktor.

Es gilt also $Z' = Z'_0 Y$. Wäre Z'_0 kompakt, also isomorph T^m, so könnte $\mathbf{0}$ kein Adhärenzpunkt von Z'_0 sein, und Z hätte die nichttriviale Komponente Z'_0, was wegen des Zusammenhangs von Z unmöglich ist. Somit ist Z'_0 nicht kompakt, d.h. es gilt $n \geqq 1$. Ist $n+m>1$, so ist $K' = K - \{\mathbf{0}\}$ zusammenhängend, d.h. es ist $Z' = Z'_0$, woraus $Z_1 \cong R^n$ und $Z_2 \cong T^m$ folgt. Von nun an sei $m < \infty$. Z' ist dann lokal einfach zusammenhängend, also ist Z einfach zusammenhängend. Ist $n+m<2$, so kann die Entfernung des Punktes $\mathbf{0}$ von Z den einfachen Zusammenhang des lokaleuklidischen Raumes nicht stören. Da dann $Z_1 Z_2$ einfach zusammen hängend sein muß, folgt $m=0$. Nun hat der zu R^n homöomorphe Raum Z' eine eindeutig bestimmte, zur Sphäre S^n homöomorphe Einpunktkompaktifizierung. Andrerseits läßt sich diese aus $\overline{Z} = Z' \cup \{\mathbf{0}\} \cup \{a\}$ durch Identifizierung von $\mathbf{0}$ und a gewinnen.

Es sei etwa R die Äquivalenzrelation auf $\overline{Z}$, deren Klassen $R(x)$ nur x umfassen, wenn $x \neq \mathbf{0}$, $x \neq a$, und deren Klasse $R(x)$ aus $\mathbf{0}$ und a besteht, wenn $x = \mathbf{0}$ oder $x = a$. Es gibt eine Kugel V_1 um $\mathbf{0}$ und eine Kugel V_2 um a in $\overline{Z}$ mit $V_1 \cap V_2 = \emptyset$. Die Menge $R(V_1 \cup V_2)$ ist eine Umgebung von $R(\mathbf{0})$ in $\overline{Z}/R$.

Nach Entfernung von $R(\mathbf{0})$ aus $R(V_1 \cup V_2)$ zerfällt $R(V_1 \cup V_2)$ in die getrennten aber selbst nicht notwendig zusammenhängenden Räume $R(V_1 - \{\mathbf{0}\})$ und $R(V_2 - \{a\})$. Da $\overline{Z}/R$ zu S^n homöomorph ist, enthält $R(V_1 \cup V_2)$ eine Kugelumgebung von $R(\mathbf{0})$, und man kann o.B.d.A. annehmen, daß schon $R(V_1 \cup V_2)$ eine solche ist. Daß sie nach Entfernung von $R(0)$ zerfällt, ist nur bei $n=1$ möglich. Da dann aber auch $R(V_1 - \{\mathbf{0}\})$ und $R(V_2 - \{\mathbf{0}\})$ in je zwei Komponenten zerfallen, kann $R(V_1 \cup V_2)$ nicht zu einem Intervall homöomorph sein. Dieser Widerspruch zeigt, daß $m \geqq 1$ und damit $n+m \leqq 2$ sein muß. Aus $n+m>1$ folgt also die Isomorphie von Z' zu $R \times T$, d.h. zur multiplikativen Gruppe der komplexen Zahlen.

Ist nun schließlich $m=0$, so ist $Z'_0 = R$. Der lokaleuklidische eindimensionale Raum Z zerfällt nach Entfernung von $\mathbf{0}$ in die Nebenklassen von Z'_0 homöomorph zu R, d.h. zu einem offenen Intervall. Der Punkt $\mathbf{0}$ besitzt aber wie jeder andere eine Umgebungsbasis aus Intervallen, die alle nach Entfernung von $\mathbf{0}$ in genau zwei Intervalle zerfallen, jedes derselben liegt aber in genau einer Nebenklasse nach Z'_0. Andrerseits hat jede Z'_0-Nebenklasse mit jeder Umgebung von $\mathbf{0}$ nichtleeren Durchschnitt, sonst wäre ja Z nicht zusammenhängend. Also hat Z' genau zwei Nebenklassen nach $Z'_0 \cong R$ und ist als abelsche Gruppe isomorph zur multiplikativen Gruppe der reellen Zahlen.

Wir haben folgenden Satz bewiesen

(3.2) *Eine lokalkompakte zusammenhängende endlichdimensionale topologische Doppelloop mit abelscher assoziativer Multiplikation hat eine zur multiplikativen Gruppe aller reellen Zahlen oder aber aller komplexen Zahlen isomorphe Multiplikationsloop.*

Nun sei $(Z, +, \cdot)$ total unzusammenhängend, aber nicht diskret. Dann ist Z_2, die größte kompakte Untergruppe von Z' offen in Z', da eine total unzusammenhängende lokalkompakte Gruppe kompakte offene Untergruppen enthält. Z'/Z_2 ist diskret, also ist Z_1 diskret und daher abgeschlossen in Z'. Der Teilraum $(A_a \cup B) \cap Z$ ist offen in Z; also ist $A_0 \cap Z$ eine kompakte offene Umgebung von 0 in Z. Die Menge $Z \cap (A_0 \cup B)$ nennen wir Q. Auf der Gruppe Z_1 wird eine Vollordnung $\prec$ definiert, wenn festgesetzt wird, daß $z \prec z'$ genau dann gelte, wenn $z' z^{-1}$ in A_0 liegt. Da $z'' \in A_0$ und $z''^{-1} \in A_a$ gleichwertig sind, da ferner $Z_1 \subset A_0 \cup A_a \cup \{1\}$ gilt, und $\prec$ transitiv ist, wird in der Tat damit eine Vollordnung definiert. Für alle $x \in Z_1$ folgt aus $z \prec z'$ auch $zx \prec z'x$. Es sei nun $z \leqq z'$ mit $1 \leqq z'$ und $1 \prec z$. Da z in A_0 liegt, gilt $\lim_{n \to \infty} z^n = 0$, also auch $\lim_{n \to \infty} z^n z'^{-1} = 0$, so daß sicher eine natürliche Zahl m existiert, für welche $z^m z'^{-1} \in A_0$, d. h. $z' \prec z^m$ richtig ist. Z_1 ist also eine archimedisch angeordnete abelsche Gruppe und daher isomorph zu einer Untergruppe der additiven Gruppe der reellen Zahlen. Nun sei z ein Element aus $A_0 \cap Z_1$. In jedem Fall gilt dann $Qz \subseteqq A_0 \cap Z$. Es sei zunächst $Qz \subset A_0 \cap Z$; dann existiert ein Element $z' \in A_0 \cap Z_1$, welches nicht in Qz liegt. Es liege nämlich etwa xz' mit $z' \in Z_1$ und $x \in Z_2$ in $A_0 \cap Z$, aber nicht in Qz; nun kann z' nicht 1 sein wegen $xz' \in A_0$; infolge $QQ \subseteqq Q$ läge mit z' auch xz' in Qz; weswegen $z' \notin Qz$ folgt. Es ist dann $z' \prec z$. Der Raum $M = (A_0 \cap Z) - Qz$ ist kompakt; also ist $Z_1 \cap M$ wegen der Abgeschlossenheit von Z_1 in Z kompakt, zugleich aber diskret, also endlich. $Z_1 \cap M$ enthält also ein $\prec$-kleinstes Element π. Mit diesem gilt dann notwendig $Q\pi = A_0 \cap Z$. Gäbe es in $Z_1 \cap A_0$ noch ein kleineres Element z_0 als π, so wäre $z_0 \in Z_1 \cap Qz$, d. h. es wäre $z_0 z^{-1} \in Q$ oder $z \prec z_0$. Da andrerseits $\pi \prec z$ gilt, ist dies unmöglich, die geordnete Gruppe Z_1 hat also ein kleinstes Element oberhalb 1 und ist dann wegen der Isomorphie zu einer additiven Gruppe ganzer Zahlen zyklisch und von π erzeugt.

Die kompakte Gruppe Z_2 ist total unzusammenhängend. Ihre Charaktergruppe ist eine diskrete Torsionsgruppe, die in primäre direkte Faktoren zerfällt. Deren duale Untergruppen in Z_2 sind Moduln über den ganzen p-adischen Zahlen ohne Elemente unendlicher Höhe, die in der p-adischen Topologie vollständig sind. Die p-adische Topologie kann feiner sein als die von der gegebenen induzierte. Diese Moduln sind die Vervollständigung einer direkten Summe zyklischer Moduln in der p-adischen Topologie (KAPLANSKY [8], S. 55).

Damit haben wir folgenden Satz bewiesen:

(3.3) *Eine lokalkompakte total unzusammenhängende nichtdiskrete topologische Doppelloop mit abelscher assoziativer Multiplikation hat eine Multiplikationsloop, die isomorph ist zu einem direkten Produkt einer unendlichen zyklischen diskreten Gruppe $\langle \pi \rangle$ und kompakten Faktoren H_p von denen jeder algebraisch isomorph ist der Vervollständigung einer direkten Summe zyklischer Moduln in der p-adischen Topologie, die nicht gröber ist als die induzierte. Das Element $\pi \in A_0$ ist modulo B eindeutig bestimmt.*

In (1.10) ist der Begriff eines K-Fastvektorraums eingeführt worden. $(G, +)$ sei ein $(K, +, \cdot)$-Fastvektorraum; $(G, +)$ heißt *topologisch*, wenn $(G, +)$ eine topologische Loop und $(K, +, \cdot)$ ein nichtdiskreter topologischer Neokörper ist, und wenn die Abbildung $(r, x) \to r \cdot x$ von $K \times G$ in G stetig ist.

(3.4) *$(G, +)$ sei ein topologischer $(K, +, \cdot)$-Fastvektorraum. K sei zusammenhängend. Dann ist G zusammenhängend.*

Beweis. Jedes Element $x \in G$ liegt in einem zusammenhängenden Unterraum $K \cdot x$ von G, der das Nullelement $\mathbf{0} \in G$ enthält.

(3.5) *$(G, +)$ sei ein topologischer lokalkompakter $(K, +, \cdot)$-Fastvektorraum. G sei zusammenhängend. Dann ist K zusammenhängend.*

Beweis. Es sei K total unzusammenhängend. g_m seien die in (2.3) eingeführten Funktionen für $(K, +)$; $\tilde{g}_m$ die entsprechenden für G. Dann gibt es zu jeder kompakten offenen Umgebung V von $\mathbf{0}$ in $(K, +)$ eine kompakte offene Umgebung W von $\mathbf{0}$ in $(K, +)$ derart, daß $g_m(W, \ldots, W) \subseteq V$ für alle natürlichen m. Ist U eine offene relativ kompakte Umgebung von $\mathbf{0}$ in G, so ist auch $W \cdot U$ eine relativ kompakte offene Umgebung von $\mathbf{0}$ in G. Für alle $x = r \cdot u \in W \cdot U$ mit $r \in W$, $u \in U$ gilt $\tilde{g}_m(x, \ldots, x) = \tilde{g}_m(r \cdot u, \ldots, r \cdot u) = \tilde{g}_m(r, \ldots, r) \cdot u \in V \cdot U$ für $m = 1, 2, \ldots$. Andrerseits ist $G = \bigcup_{m=1}^{\infty} \tilde{g}_m(W \cdot U, \ldots, W \cdot U)$, da G zusammenhängend ist. Es gibt also wegen der Kompaktheit von $V\overline{U}$ ein n, so daß $V \cdot U \subseteq \tilde{g}_{n-1}(W \cdot U, \ldots, W \cdot U)$. Nach Hilfssatz (2.3) finden wir nun in $W \cdot U$ ein x derart, daß $g_n(x, \ldots, x) \notin \overline{g_{n-1}(W \cdot U, \ldots, W \cdot U)}$; erst recht ist dann $\tilde{g}_n(x, \ldots, x) \notin V \cdot U$. Dieser Widerspruch zeigt, daß K zusammenhängend sein muß.

Damit können wir folgenden Satz beweisen:

(3.6) *Ein lokalkompakter topologischer K-Fastvektorraum ist entweder zusammenhängend oder total unzusammenhängend. Er ist genau dann zusammenhängend, wenn K zusammenhängend ist.*

Beweis. Zum Beweis fehlt uns noch, daß ein lokalkompakter K-Fastvektorraum über einem total unzusammenhängenden Neokörper K total unzusammenhängend ist. Es sei H die Zusammenhangskomponente von $\mathbf{0}$ in G. Wir nehmen an, sie bestehe aus mehr als einem Punkt und sei also nicht diskret. Dann ist $K \cdot H$ Vereinigung zusammenhängender Unterräume, die alle $\mathbf{0} \in G$ enthalten; diese ist als solche in H enthalten. H ist also ein zusammenhängender K-Fastvektorraum, wenn H nichtdiskret ist. Dann müßte nach (3.4) aber K zusammenhängend sein. Also ist $H = \mathbf{0}$, w.z.b.w.

Wir fassen unsere Kenntnisse nun für distributive Doppelloops in folgendem Satz zusammen:

SATZ I

(3.7) *$(K, +, \cdot)$ sei eine distributive lokalkompakte endlichdimensionale topologische Doppelloop. $(Z, +, \cdot)$ sei ihr Zentrum. Wir betrachten folgende Aussagen.*

a) *K ist zusammenhängend.*

b) *K ist total unzusammenhängend und habe ein nichtdiskretes Zentrum.*

A) $(Z', \cdot)$ *ist isomorph zur multiplikativen Gruppe aller reellen Zahlen oder aber zur multiplikativen Gruppe aller komplexen Zahlen.*

B) $(Z', \cdot)$ *ist isomorph zu einem direkten Produkt aus einer unendlichen diskreten zyklischen Gruppe* $\langle \pi \rangle$ *und einer eindeutig bestimmten kompakten total unzusammenhängenden abelschen Gruppe, die ihrerseits direktes Produkt ist von Faktoren* H_p, *die als abstrakte Gruppen p-adische Vervollständigung von direkten Produkten zyklischer Moduln sind.*

Dann gelten folgende Bezeichnungen:

a) *ist gleichwertig zu* A),

b) *ist gleichwertig zu* B).

Beweis. Die beiden Behauptungen folgen aus (1.11), (3.5), (3.2) und (3.3). Wir haben nur noch nachzuweisen, daß Z im zusammenhängenden Fall nicht diskret ist. Wäre Z diskret, so gäbe es eine Umgebung $V \in \mathfrak{U}(\mathbf{0})$ in $\bar{K}$ die mit Z den Durchschnitt $\{\mathbf{0}\}$ hätte. Dann hätte $\bar{Z}$ mit $f(V)$ den Durchschnitt $\{a\}$, d.h. Z wäre kompakt, also endlich (Definitionen wie am Anfang von § 3). $(K, +)$ kann dann aber kein Z-Fastvektorraum sein, wie aus dem Beweis von (3.5) ersichtlich wird, wenn man $W \cdot U$ und $V \cdot U$ durch $U \cdot Z$ ersetzt.

Wir spezialisieren nun noch die Voraussetzung (3.7) a) und können in einem Fall sogar auf ein Distributivgesetz verzichten. Wir nennen einen topologischen *Neokörper reell* bzw. *komplex*, wenn $(K', \cdot)$ zur multiplikativen Gruppe der reellen bzw. komplexen Zahlen isomorph ist. Die folgenden Sätze geben Kennzeichnungen reeller und komplexer Neokörper, die nach (3.7) als Zentren zusammenhängender Doppelloops auftreten.

Satz II

(3.8) $(K, +, \cdot)$ *sei eine rechtsdistributive topologische Doppelloop, die lokal einem Intervall der reellen Geraden homöomorph sei. Dann ist* $(K, +, \cdot)$ *ein reeller Neokörper.*

Beweis. H' sei der Rechtskern von K' [vgl. (1.5)]. Wir beweisen, daß H' nichtdiskret ist; dann enthält jede Umgebung von **1** ein Element von H'; und K' ist monassoziativ konzentriert im Sinne von [4], (1.2). K' enthält dann eine Einparametergruppe E [ebenda II, (6.4) c)], die ein Intervall um **1**, genauer die Komponente von **1** in K' überdeckt. H' häuft sich also in E gegen **1** und ist daher in E dicht, also in einem Intervall um **1**. Da H eine abgeschlossene Unterdoppelloop von K ist, folgt daraus dann $H = K$. Dann ist aber $K' = H'$ isomorph zur multiplikativen Gruppe der reellen Zahlen (vgl. [4], Beweis von (2.3)).

Angenommen, H' wäre diskret; dann wäre H' kompakt, denn eine Umgebung V von **0** ist frei von Elementen von H', dann aber auch $f(V)$ (Definition von f im ersten Absatz von § 3). Nun ist also H' endlich und enthält ein von **1** verschiedenes Element z, dessen sämtliche Potenzen in einem kompakten Intervall um **1** liegen. Dies ist aber unmöglich, wie der Verfasser in [4], (6.1) gezeigt hat, oder wie man aus Hilfssatz (3.10) erkennen wird.

(3.9) *$(K, +, \cdot)$ sei eine rechtsdistributive topologische Doppelloop, die lokal einer Kreisscheibe homöomorph sei. Dann ist $(K, +, \cdot)$ ein komplexer Neokörper oder enthält als Rechtskern einen reellen Neokörper.*

Beweis. $(H, +, \cdot)$ sei der Rechtskern von $(K, +, \cdot)$. Nun ist K zwar kein H-Fastvektorraum, denn die Linksmultiplikationen mit Elementen aus H sind im allgemeinen keine Automorphismen von $(K, +)$. Es gilt jedoch $(r+s)\cdot x = r\cdot x + s\cdot x$ für r und s aus H. Damit ist der Beweis von (3.5) durchführbar, der zeigt, daß H nicht total unzusammenhängend und dabei nicht diskret sein kann. Ebenso wie im Beweis von (3.7) sieht man auch, daß H nicht diskret ist. H ist also eine zusammenhängende Unterdoppelloop von K. Demnach ist H' eine zusammenhängende lokalkompakte nicht kompakte Gruppe der Dimension $\leqq 2$. Ist $\dim H = 1$, so sind wir im Fall von (3.8) und $(H, +\cdot)$ ist ein reeller Neokörper. Ist $\dim H = 2$, so ist $K = H$ und H' ist eine Liegruppe, die zu der Zylindergruppe $R\times T$ homöomorph ist (R = Gruppe der reellen Zahlen, T = Torusgruppe).

Die einzige zweidimensionale Liegruppe, für die das der Fall ist, ist aber isomorph zu $R\times T$, also zur multiplikativen Gruppe der reellen Zahlen.

Hilfssatz:

(3.10) *E sei ein homogener separabler zusammenhängender topologischer Raum, der einem Jordanbogen lokal homöomorph sei. $x\to h(x)$ sei ein Homöomorphismus von E der keinen Punkt festläßt. Ein Punkt $p\in E$ hat eine Umgebung U, die einem kompakten Intervall homöomorph ist. Dann gibt es eine natürliche Zahl n, so daß $h^n(p) = h\big(h(\ldots(h(p))\ldots)$ nicht in U liegt.*

Beweis. (R, f) sei der einfach zusammenhängende Überlagerungsraum; R ist bis auf Homöomorphie der Raum der reellen Zahlen, f sei die Überlagerungsabbildung $R\to E$. Zu der Abbildung $h\circ f$ von R in E gibt es eine Abbildung $\tilde{h}$ von R auf R, für welche $f\circ\tilde{h} = h\circ f$ gilt; schreiben wir für einen Punkt $r\in R$ das Bild $\tilde{h}(r)$ in $f^{-1}\circ h\circ f(r)$ vor, so ist $\tilde{h}$ eindeutig bestimmt (CHEVALLEY [2]). $\tilde{h}$ ist ein lokaler Homöomorphismus für jeden Punkt von R, d.h. zu jedem $r\in R$ gibt es ein Intervall U_r, auf dem $\tilde{h}$ echt monoton ist; man braucht ja nur ein Intervall U_r zu suchen, so daß $f(U_r)$ schlicht überlagert wird.

Dann ist aber $\tilde{h}$ im Großen monoton; sei nämlich V_r die Vereinigung aller Intervalle U_r um r, auf denen h monoton ist. Dann ist V_r einerseits offen, da h lokal monoton ist, und abgeschlossen andrerseits; denn ist $s\in\overline{V_r}$, so ist $U_s\cap V_r = \emptyset$ und h ist auch noch auf $U_s\cup V_r$ monoton, d.h. $s\in U_s\subseteqq V_r$. Es gilt also $V_r = R$. Wir können uns beim Beweis der Behauptung also darauf beschränken, daß E zu R homöomorph sei. h ist echt monoton wachsend; andernfalls hätte h einen Fixpunkt. Die Folge $\big(h^n(x)\big)_{n=1,\ldots}$ ist dann monoton. Ist nämlich $h(x)\gtrless x$, so folgt $h\big(h(x)\big)\gtrless h(x)$ und $h^n(x)\gtrless h^{n-1}(x)$ für alle $n>1$. Lägen nun alle $h^n(x)$ in U, so würde $\lim\limits_{n\to\infty} h^n(x) = y$ existieren und es wäre $h(y) = \lim\limits_{n\to\infty} h^{n+1}(x) = y$ ein Fixpunkt im Widerspruch zur Voraussetzung.

(3.11) *$(K, +, \cdot)$ sei eine rechtsdistributive topologische Doppelloop, die lokal einer Kreisscheibe homöomorph sei. $(K', \cdot)$ sei monassoziativ. Dann enthält $(K', \cdot)$ wenigstens eine zur additiven reellen Gruppe isomorphe Einparametergruppe L und eine davon verschiedene Einparameterhalbgruppe E; das Produkt dieser zwei Untergruppen ist K'. Ist die Halbgruppe E kompakt, so ist sie einer Torusgruppe isomorph.*

Beweis. $(H, +, \cdot)$ sei der Rechtskern von K. Der Fall $H = K$ ist trivial. Es sei $H \neq K$. Die Zusammenhangskomponente von $\mathbf{1}$ in H heiße L. Jedes Element $x \in K'$ liegt genau in einer Menge der Form Lx; denn ist $l_1 x \in Ly$ mit $l_1 \in L$, so folgt $x \in l_1^{-1}(Ly) = Ly$ und der Symmetrie wegen auch $y \in Lx$. Durch L wird also eine abgeschlossene Äquivalenzrelation erklärt, die wir mit l bezeichnen wollen; es gelte $x\,l\,y$ genau dann, wenn $x \in Ly$. Auf dem Restklassenraum K'/l ist durch die Abbildung $h = (Ly \to (Lx)y)$ mit $x \in K' - L$ ein Homöomorphismus ohne Fixpunkt erklärt; denn h ist eine stetige und umkehrbar stetige Abbildung K'/l auf sich. Mit Hilfe der an anderer Stelle ([5], (4.7)) entwickelten Methoden erkennt man, daß K'/l einer Jordankurve homöomorph ist. U sei eine kompakte Umgebung von $\mathbf{1}$ in K' derart, daß $l(U)$ in K'/l einem Intervall homöomorph ist. Dann gibt es nach Hilfssatz (3.10) eine Zahl n so, daß für $x \in K' - L$ allemal $h^n(L) \in l(U)$ oder $(Lx^n) = (\ldots(Lx)x \ldots)x \notin L \cdot U$ gilt. Insbesondere ist $x^n \notin U$. Ist $\subset L$, $x \neq \mathbf{1}$, so gibt es auch eine natürliche Zahl n so, daß $x^n \notin U$. Die Loop $(K',)$ ist also eudoxisch im Sinne von [4] (3.1). Nun sei $(x_i)_{i=1,\ldots}$ eine gegen $\mathbf{1}$ konvergierende Folge von Elementen aus U, die nicht in L liegen; mit m_i sei die kleinste natürliche Zahl bezeichnet, für die noch $L x_i^{m_i+1} \not\subseteq LU$ gilt.

Für alle reellen $r \in [0, 1]$ sei $[r m_i]$ die kleinste ganze Zahl nicht oberhalb $r m_i$. Es gibt eine Teilfolge $(i_l)_{l=1,\ldots}$ der natürlichen Zahlen, so daß $\varphi(r) = \lim_{l \to \infty} x_{i_l}^{[r m_{i_l}]}$ existiert und das Intervall $[0, 1]$ homöomorph in LU abbildet, so daß $\varphi(r_1 + r_2) = \varphi(r_1)\,\varphi(r_2)$ gilt sobald r_1, r_2 und $r_1 + r_2$ in $[0, 1]$ liegen [Hofmann, II, (5.1)]. $\varphi(1)$ liegt auf dem Rand von LU. Nun ist $l(\varphi(1)) = \lim_{l \to \infty} l(x_{i_l}^{m_{i_l}}) = \lim_{l \to \infty} l(x_{i_l}^{m_{i_l}+1})$ wegen $x_{i_l} \to \mathbf{1}$. Da $l(x_{i_l}^{m_{i_l}}) \in LU$ und $l(x_{i_l}^{m_{i_l}+1}) \notin LU$, ist $l(\varphi(1))$ auf dem Rand von $l(U)$ gelegen. $\varphi(1)$ liegt also sicher nicht in L. Die Abbildung läßt sich eindeutig fortsetzen zu einer stetigen algebraisch homomorphen Abbildung φ der additiven Halbgruppe R_+ aller nichtnegativen reellen Zahlen. $E = \varphi(R_+)$ ist eine Einparameterhalbgruppe. Es gibt eine kleinste reelle Zahl $r > 0$ derart, daß $l(\varphi(r)) = L$, d.h. daß $\varphi(r) \in L$; wir können ohne Beschränkung der Allgemeinheit $r = 1$ annehmen. Es gilt dann $K' = L \cdot \varphi([0, 1])$.

Genau dann ist E kompakt, wenn $\varphi(1) = \mathbf{1}$. Dann ist E isomorph einer Torusgruppe T: [4], (6.4).

(3.12) Die Voraussetzungen seien dieselben wie in (3.11); aber K sei sogar distributiv. Dann ist K ein komplexer Neokörper.

Beweis. $(H, +, \cdot)$ sei das Zentrum von K. Im Fall $H = K$ ist nichts zu beweisen. Sei also H ein reeller Neokörper. Wir konstruieren φ wie in (3.11). ψ sei ein Isomorphismus der additiven Gruppe R der reellen Zahlen auf L,

die Komponente von 1 in H'. O.B.d.A. sei $\psi(1)=\varphi(1)$. Die Abbildung $r\to\varphi(r)\,\psi(r)^{-1}$ ist ein stetiger algebraischer Homomorphismus von R in K' mit $\varphi(1)\,\psi(1)^{-1}=1$ und $\varphi(r)\,\psi(r)^{-1}\neq 1$ für $0<r<1$. Das Bild von R bei diesem Homomorphismus ist eine zu einer Torusgruppe isomorphe Einparametergruppe E. Da L zentral ist, ist K' das direkte Produkt von E und L.

(3.13) Die Voraussetzungen seien dieselben wie in (3.11); aber $(K',\cdot)$ sei sogar potenzassoziativ. Dann kann E kompakt bestimmt werden, und $(K,+,\cdot)$ ist ein komplexer Neokörper, wenn $(K',\cdot)$ auch noch diassoziativ ist.

Beweis. L sei bestimmt wie in (3.11). Die Teilräume A_0 und A_a sind in der 2-Sphäre $\bar{K}=K\cup\{a\}$ offen und punktfremd. Das Komplement B ist dann ein kompakter zusammenhängender Teilraum von $\bar{K}$, der alle kompakten Unterloops von $(K',\cdot)$ enthält. Wir wählen dann die Folge $(x_i)_{i=1,\dots}$ von (3.11) in B; das ist möglich, da B das Element 1 enthält und zusammenhängend ist. Dann liegt die durch φ definierte Einparametergruppe E in B; sie ist also relativ kompakt. Da $(K',\cdot)$ zweidimensional und einem Zylinder homöomorph ist, kann E nicht in einer echt größeren zusammenhängenden abelschen Gruppe dicht sein. Sie ist also eine eindimensionale Torusgruppe. Ist nun $(K',\cdot)$ alternativ, so liegen L und E in einer abgeschlossenen, zusammenhängenden, mindestens zweidimensionalen Gruppe G, wie man leicht sieht, wenn man die Hülle der Vereinigung aller von $\varphi(2^{-n})$ und $\psi(2^{-n})$ erzeugten Gruppen betrachtet. Da $(K',\cdot)$ zweidimensional ist, ist das Erzeugnis von L und E genau zweidimensional. Sie muß dann einem Zylinder homöomorph sein und stellt dann sogar eine Zylindergruppe dar. G ist offen in K', da G und K' beide zweidimensional sind. Dann ist G aber auch abgeschlossen in K' ([3] (3.1)) und füllt daher den zusammenhängenden Raum K' ganz aus.

Wir fassen unsere Ergebnisse in folgendem Satz zusammen:

SATZ III

(3.14) *$(K,+,\cdot)$ sei eine topologische Doppelloop, die lokal einer Kreisscheibe homöomorph sei. Wir betrachten folgende Aussagen*

a) *$(K,+,\cdot)$ ist rechtsdistributiv,*

b) *$(K,+,\cdot)$ ist distributiv,*

c) *$(K',\cdot)$ ist monassoziativ,*

d) *$(K',\cdot)$ ist diassoziativ*

A) *$(K,+,\cdot)$ ist ein komplexer Neokörper,*

B) *$(K,+,\cdot)$ enthält einen reellen Neokörper als Rechtskern.*

Dann gelten folgende Beziehungen:

Aus a) *folgt* A) *oder* B).

Aus b) *und* c) *folgt* A).

Aus a) *und* d) *folgt* A).

Man beachte, daß A) nicht B) nach sich zieht [vgl. S. 67, Absatz 3].

Wir wenden uns der Struktur allgemeiner distributiver Doppelloops zu und behandeln zunächst den zusammenhängenden Fall, der sich besser übersehen läßt als der total unzusammenhängende.

(3.15) $(K, +, \cdot)$ sei eine lokalkompakte zusammenhängende distributive topologische Doppelloop mit potenzassoziativer Multiplikation. Mit L bezeichnen wir eine nichtkompakte zentrale Einparameteruntergruppe von $(K', \cdot)$, die nach (3.7) in $(Z', \cdot)$ existiert. Es gibt dann stetige Abbildungen $x \to l(x)$ von K' auf L und $x \to b(x)$ von K' auf B, derart, daß $x = l(x) \cdot b(x)$; es ist demnach $L \cdot B$ topologisch ist also K' homöomorph dem Produktraum $L \times B$. Bezeichnet h den Restklassenhomomorphismus von K' auf K'/L, so wird B durch h homöomorph auf K'/L abgebildet. Jede kompakte Unterloop von K' ist in B enthalten.

Beweis. Für jedes $x \in K'$ ist xL ein zusammenhängender offener Jordanbogen mit den Adhärenzpunkten $\mathbf{0}$ und a in $\overline{K}$, denn $\lim_{n \to \infty} xz^n = \mathbf{0}$ für $z \in L \cap A_0$. Sowohl $A_0 \cap xL$ wie $A_a \cap xL$ sind offene Teile von xL, da A_0 und A_a in K' offen sind. Es ist also sicher $xL \cap B \neq \emptyset$. Sind nun xl_1 und xl_2 Punkte von B mit $(l_1, l_2) \in L \times L$, so folgt, da l_1 und l_2 im Zentrum liegen, $(xl_1)(l_1^{-1}l_2) = xl_2$. Da $\overline{\langle xl_2 \rangle}$ und $\overline{\langle xl_1 \rangle}$ kompakt sind, muß auch $\overline{\langle l_1^{-1}l_2 \rangle}$ kompakt sein. Dann ist aber $l_1^{-1}l_2 = \mathbf{1}$. Es gilt demnach $K' = L \cdot B$ und h bildet den kompakten Raum eineindeutig stetig, also homöomorph auf K'/L ab. $l(x)$ und $b(x)$ definieren wir durch die eindeutige Zerlegung $x = l(x) \cdot b(x)$. Da bei $x \to b(x)$ offene Teilräume von B offene Urbilder haben sind b und damit aber auch l stetig. Da jedes Element einer kompakten Unterloop von K' eine relativ kompakte Untergruppe erzeugt, ist jede kompakte Unterloop in B enthalten.

Wir bezeichnen im folgenden einen topologischen Neokörper, dessen multiplikative Gruppe der multiplikativen Gruppe des Quaternionenschiefkörpers über dem Körper der reellen Zahlen isomorph ist, als einen *quaternalen Neokörper*.

(3.16) Zu den Voraussetzungen von (3.15) komme noch die Forderung hinzu, $(K', \cdot)$ möge assoziativ sein. Dann ist $(K', \cdot)$ isomorph dem Produkt einer Torusgruppe mit einem projektiven Limes kompakter und im Fall eines endlichdimensionalen Zentrums einfacher Liegruppen. Ist K endlichdimensional, dann ist $(K, +, \cdot)$ ein reeller, ein komplexer oder ein quaternaler Neokörper.

Beweis. O.B.d.A. nehmen wir $\dim K' > 2$ an, da der zweidimensionale Fall mit Satz III erledigt ist. K' ist dann eine zusammenhängende lokalkompakte Gruppe mit einer zentralen Untergruppe der Form $L \cdot D$, wo L eine nichtkompakte, zur additiven reellen Gruppe $(R, +)$ isomorphe Einparametergruppe und D eine Gruppe aus zwei Elementen bezeichnet (Satz I). Es gilt $K' = L \cdot B$ und B ist die Vereinigung aller maximalen kompakten und zusammenhängenden Untergruppen von K' (vgl. z.B. MONTGOMERY-ZIPPIN [9], S. 188, Satz v. IWASAWA). Es sei M eine solche. Dann existieren in K nicht kompakte Einparametergruppen $E_1, \ldots, E_n$ derart, daß die Abbildung

$(l_1, \dots, l_n, m) \to l_1 \dots l_n m$ des Produktraums $E_1 \times \cdots \times E_n \times M$ in K' einen Homöomorphismus auf K' darstellt.

Es ist nun klar, daß man eine normale nichtkompakte Einparametergruppe E unter die $E_1, \dots, E_n$ aufnehmen kann, da man die Faktorgruppe nach E wieder in der angegebenen Weise in Einparametergruppe $\bar{E}_m$ und eine zu M isomorphe maximale kompakte Gruppe zerlegen und in jeder Einparametergruppe $\bar{E}_m$ der Faktorgruppe eine Vertretergruppe E'_m wählen kann; $E'_1, \dots, E'_{m-1}, E$ liefert dann einen Satz Einparametergruppen der gewünschten Art. Wir nehmen daher o.B.d.A. $E_1 = L$ an. Da aber K'/L kompakt ist, ist $n = 1$ und $L \times M$ wird durch $(l, m) \to l \cdot m$ homöomorph auf $K' = L \cdot B$ abgebildet; $M \subseteq B$ ist isomorph K'/L bei der Projektion h von K' auf K'/L. Bei dieser wird aber B eineindeutig abgebildet. Daher ist $M = B$ und K' ist direktes Produkt $L \cdot B$ aus einer zentralen nichtkompakten Einparametergruppe L und einer größten zusammenhängenden und bei $\infty > \dim K \geqq 2$ einfach zusammenhängenden Gruppe B. Angenommen nun Z sei zweidimensional; nach Satz I kann ja Z bei $\dim Z < \infty$ nur ein- oder zweidimensional sein. Dann hat das Zentrum von B eine eindimensionale Torusgruppe T als Zusammenhangskomponente von $\mathbf{1}$. Nun ist aber B Produkt AH aus einer zusammenhängenden kompakten abelschen zentralen Gruppe und einem projektiven Limes H halbeinfacher Liegruppen (vgl. z. B. WEIL [17], S. 89–90). In unserem Fall gilt $A = T$. Der Durchschnitt $T \cap H$ ist abgeschlossen und ungleich T, also endlich. B ist isomorph einer Faktorgruppe des direkten Produkts $T \times H$ nach einem diskreten zentralen Normalteiler (A. WEIL [17], S. 21). Da $T \times H$ nicht einfach zusammenhängend ist, gilt dies erst recht für die Faktorgruppe, die aber im Fall $\dim K < \infty$ andrerseits einfach zusammenhängen muß. Die Annahme $\dim Z = 2$ hat also in diesem Fall zum Widerspruch geführt. Das Zentrum von B besteht dann also gerade aus D, einer Gruppe aus zwei Elementen und B ist ein direktes Produkt einfacher und einfach zusammenhängender Liegruppen (WEIL [17], S. 91). Ist die Dimension von K endlich, so besteht dieses Produkt nur aus endlich vielen Faktoren und ist eine Liegruppe. Die Mannigfaltigkeit $((L \cap A_0) \cup \{\mathbf{0}\})\, B$ ist eine topologische Halbgruppe mit dem regulären Rand B und $\mathbf{0}$ als Null und $\mathbf{1}$ als Eins. $\mathbf{0}$ bleibt bei Multiplikation mit allen Elementen, auch insbesondere mit Elementen aus B fest. Nach einem Satz aus [10], MOSTERT und SHIELDS, S. 118 ist B eine 0-, 1- oder 3-Sphäre. Dabei wird von diesem Satz nur der Teil benötigt, der von MOSTERT in [11] bewiesen wurde. Da B mindestens dreidimensional sein muß, ist B homöomorph einer 3-Sphäre. Dann ist B notwendig isomorph der einfach zusammenhängenden Überlagerungsgruppe der speziellen orthogonalen Gruppe $SO(3)$. Sie ist also eindeutig bestimmt und somit auch isomorph zu der multiplikativen Gruppe des Quaternionenschiefkörpers, der ja ein spezieller Neokörper ist. Damit ist der Beweis vollendet.

(3.17) $(K, +\,\cdot)$ sei eine total unzusammenhängende lokalkompakte topologische distributive Doppelloop mit potenzassoziativer Multiplikation und nichtdiskretem Zentrum. L sei die in (3.3) eingeführte unendliche diskrete zyklische

Untergruppe $\langle \pi \rangle = Z_1$. Dann ist K'/L kompakt. B ist eine Umgebung des neutralen Elements, die jede kompakte Unterloop von K' enthält. Bezeichnet h wieder den Restklassenhomomorphismus von K' auf K'/L, so wird B durch h homöomorph in K'/L abgebildet und das Bild $h(B)$ ist eine Umgebung von $h(1)$.

Beweis. Zunächst steht fest, daß K'/Z' kompakt ist, denn K ist ein Z-Fastvektorraum und alle Zx mit $x \in K'$ schneiden den kompakten Teilraum $V-W$, wo V eine kompakte und W eine offene Umgebung von 0 bezeichnen, die in der Beziehung $W+W \subset V$ zueinander stehen. Nun ist $Z' = Z_1 \cdot Z_2 = L \cdot Z_2$, und wenn $x \cdot L \cdot Z_2 \cap (V-W) \neq \emptyset$ gilt, dann gilt auch für alle $x \in K'$ stets $x \cdot L \cap Z_2(V-W) \neq \emptyset$. Da aber $Z_2(V-W)$ immer noch kompakt ist, folgt daraus die Behauptung, daß K'/L kompakt sei. Ferner sieht man, wie im entsprechenden Stück des Beweises von (3.15), daß xL den Teilraum B in höchstens einem Punkt schneiden kann. Daß dies in einer Umgebung von 1 stets wirklich der Fall ist, erkennt man daran, daß B eine offene Umgebung V_1 enthält. Zu einer vorgegebenen kompakten Umgebung W_1 von 1 wählen wir nämlich $V_1 = V_1^{-1}$ so, daß $g_m(W_1, V_1, \ldots, V_1) \subseteq W_1$ für alle $m = 2, 3, \ldots$ gilt. Dann ist für alle $x \in V_1$ stets $\langle x \rangle$ in W_1 enthalten und $\overline{\langle x \rangle}$ ist daher kompakt. Während wir in B im zusammenhängenden Fall einen globalen Querschnitt vor uns hatten handelt es sich hier möglicherweise nur um einen lokalen Querschnitt.

Ist nun wiederum $(K', \cdot)$ assoziativ, so enthält jede Umgebung von 1 einen offenen kompakten Normalteiler (MONTGOMERY-ZIPPIN [9], S. 56). Es sei M ein maximaler in B enthaltener offener Normalteiler (ZORN!). Dann ist K'/M diskret und enthält in seinem Zentrum eine zu L isomorphe Untergruppe ML/M. Die Faktorgruppe nach ihr ist endlich. Ist H_0 ein endlicher Normalteiler von K'/M, so ist sein Urbild in K' ein kompakter Normalteiler, der M enthält; dieser stimmt wegen der Maximalität von M mit H überein; d.h. K'/M enthält keine endlichen nichttrivialen Normalteiler. Solche Gruppen nennen wir endlich einfach.

Wir beweisen nun folgenden Hilfssatz, und zwar ohne allgemeine gruppentheoretische Hilfsmittel in Anspruch zu nehmen:

(3.18) G eine endlicheinfache Gruppe mit einer zentralen unendlichen zyklischen Untergruppe $H = \langle c \rangle$ mit endlichem Index, so ist G zyklisch gleich $\langle c_0 \rangle$ mit $c_0^m = c$.

Beweis. a) Das Zentrum von G ist zyklisch. Es sei H_0 eine maximale zyklische zentrale Untergruppe, die H umfaßt. G/H_0 ist endlich. Das Zentrum ist also endlich erzeugbar und zerfällt in ein direktes Produkt zyklischer und, da G endlich einfach ist, unendlicher Faktoren. Da aber G/H_0 endlich ist, ist H_0 der einzige mögliche Faktor.

b) Bezeichnet man mit $x \to \bar{x}$ den Restklassenhomomorphismus von G auf G/H, so ist $\bar{x}\bar{y} = \bar{y}\bar{x}$ mit $xy = yx$ gleichwertig. Es sei nämlich x ein Element von G und $\bar{y}$ aus dem Zentralisator von $\bar{x}$. Dann gilt $y^{-1}xy = x \cdot z(\bar{y})$. Die Abbildung $\bar{y} \to z(\bar{y})$ ist ein Homomorphismus des endlichen Zentralisators von

$\bar{\bar{x}}$ in die torsionsfreie Gruppe H. Es muß also $z(\bar{y}) = \mathbf{1}$ sein und y liegt im Zentralisator von x.

c) $G/H = \overline{G}$ sei abelsch. Nach b) ist dann G abelsch und nach a) unendlich zyklisch. Da eine p-Gruppe ein nichtleeres Zentrum hat, folgt die Richtigkeit des Satzes für p-Gruppen durch vollständige Induktion nach $|\overline{G}|$.

d) Der Satz sei nun im allgemeinen nicht richtig. Dann gibt es ein Gegenbeispiel G, für welches $\overline{G} = G/H$ minimale, aber wegen c) zusammengesetzte Ordnung hat. Dann ist $\overline{G}$ einfach nichtabelsch. Denn wäre N ein Normalteiler von G, der echt zwischen H und G läge, so wäre der Satz für N richtig und N wäre zyklisch gleich $\langle c' \rangle$ mit $c'^{m_0} = c$. Wäre $g \in G$, so gälte $g^{-1}c'g = c'^{\pm 1}$, wegen $g^{-1}c'^{m_0}g = c'^{m_0}$ aber sogar $g^{-1}c'g = c'$, d. h. N läge im Zentrum von G; da aber $|G/N| < |G/H|$, könnte G kein Gegenbeispiel zu dem behaupteten Satz sein.

Wir suchen nun k Untergruppen $M_1, \ldots, M_k$ von G, so daß $\overline{M}_1, \ldots, \overline{M}_k$ maximale Untergruppen von $\overline{G}$ sind; solche existieren, da $\overline{G}$ zusammengesetzte Ordnung hat. Der Normalisator eines jeden $\overline{M}_n$ ist gleich $\overline{M}_n$. Falls $n \neq n'$ ist $\overline{M}_n \cap \overline{M}_{n'} = \{\bar{1}\}$; denn für alle M_n ist der Satz richtig; es ist also $M_n = \langle c_n \rangle$ mit $c_n^{m_n} = c$. Wenn $n \neq n'$ ist, wird G von c_n und $c_{n'}$ erzeugt; $M_n \cap M_{n'}$ ist dann normal in G und muß also wegen der Einfachheit von G/H mit H übereinstimmen. Vermöge der inneren Automorphismen $\bar{g} \to \bar{g}^{-1}\overline{M}_n\bar{g} = \overline{M}_{n(\bar{g})}$ wirkt $\overline{G}$ als Permutationsgruppe auf $(\overline{M}_1, \ldots, \overline{M}_k)$. Diejenige Untergruppe, die $\overline{M}_n$ festläßt, ist der Normalisator von $\overline{M}_n$ und stimmt daher mit $\overline{M}_n$ überein. Die Transitivitätsgebiete von $\overline{G}$ auf $(\overline{M}_1, \ldots, \overline{M}_k)$ seien $T_1, \ldots, T_l$ und $\overline{M}_n$ liege in $T_{s(n)}$. Dann gilt bekanntlich $|\overline{M}_n| \cdot |T_{s(n)}| = |\overline{G}|$. Da jedes Element von $\overline{G}$ in einer maximalen Untergruppe liegt, lassen sie die Elemente von $\overline{G}$ in folgender Weise abzählen:

$$|G/H| = |\overline{G}| = \sum_{1 \leq s(n) \leq l} \left(|\overline{M}_n| - 1\right) |T_{s(n)}| + 1 = l \cdot |\overline{G}| - \sum_{n=1}^{l} |T_n| + 1 = l \cdot |\overline{G}| - k + 1.$$

Somit gilt $(l-1)|\overline{G}| = k - 1$. Wegen $k < |\overline{G}|$ folgt daraus $l = 1$, $k = 1$, d. h. $\overline{M}_1 = \cdots = \overline{M}_k$ ist normal in $\overline{G}$ im Widerspruch zur Einfachheit von $\overline{G}$. Damit ist der Hilfssatz bewiesen.

Somit ist K'/M eine zyklische Gruppe $\langle c \cdot M \rangle$. Ist π' ein Element von K' aus der Restklasse $c \cdot M$, dann ist $\langle \pi' \rangle$ eine unendliche diskrete zyklische Gruppe, die mit M den Durchschnitt $\{\mathbf{1}\}$ hat. Es folgt daraus aber $B = M$ und $K' = M \cdot \langle \pi' \rangle$. Wir müssen offen lassen, ob dieses Produkt direkt sein muß.

Die bislang gewonnenen Ergebnisse formulieren wir in folgendem *Hauptsatz:*

SATZ IV

(3.19) *$(K, +, \cdot)$ sei eine topologische lokalkompakte distributive Doppelloop mit potenzassoziativer Multiplikationsloop $(K', \cdot)$.*

Dann gelten folgende Aussagen:

A) *K' enthält eine außer für* $\dim K = 0$ *und 2 eindeutig bestimmte zentrale nichtkompakte Untergruppe L derart, daß K'/L kompakt ist. Genau wenn K*

zusammenhängend ist, ist L eine Einparametergruppe, genau wenn K total unzusammenhängend ist, ist L eine unendliche zyklische Gruppe.

B) *Der kompakte Raum B aller Elemente mit nicht konvergierenden oder divergierenden positiven Potenzen wird bei dem Restklassenhomomorphismus h von K' eineindeutig in K'/L abgebildet und ist ein lokaler Querschnitt; d.h. $h(B)$ ist eine Umgebung von $h(1)$ und die Einschränkung von h auf B ist ein Homöomorphismus. Ist K zusammenhängend, so ist $K' = L \cdot B$ und B ist homöomorph K'/L.*

C) *Rechtskern, Linkskern und Kern von $(K, +, \cdot)$ sind distributive Doppelloops mit assoziativer Multiplikation*, d.h. *Neokörper $(H, +, \cdot)$ folgender Struktur: $(H', \cdot)$ ist Produkt aus einer nichtkompakten Untergruppe L_0 und einem eindeutig bestimmten maximalen kompakten Normalteiler M.*

Ist K zusammenhängend, so ist $L_0 = L$ eine zentrale Einparametergruppe und M ist ein projektiver Limes kompakter Liegruppen. Ist K endlichdimensional, so ist M eine 0-, 1- oder 3-Sphäre, d.h. $(H, +, \cdot)$ ist ein reeller, ein komplexer oder ein quaternaler Neokörper. Ist K total unzusammenhängend, so ist L_0 eine unendliche zyklische Gruppe und M ist offen, H'/LM mit der in A) *erwähnten zentralen Untergruppe L ist eine endliche zyklische Gruppe.*

§ 4. Die Addition in Neokörpern

(4.1) *Auf Grund der Rechtsdistributivität in einer distributiven Doppelloop K wird die Addition bei Kenntnis der Multiplikationsloop völlig durch die Abbildung $x \to (1 + x) = \alpha(x)$ gekennzeichnet. Es gilt nämlich $(x + y) = (1 + y/x)\, x = \alpha(y/x)\, x$* falls $x \neq 0$. *Die Abbildung α muß folgende Eigenschaften haben:*

a) *$x \to \alpha(x)$ ist eine eineindeutige Abbildung von K auf K.*

b) *$x \to \alpha(z/x)\, x$ ist eine eineindeutige Abbildung von K' auf $K - \{z\}$ für alle $z \in K'$.*

b') *Es gilt $\alpha(z/x)\, x = x\alpha(x\backslash z)$ für alle $x \in K'$ und $z \in K$.*

c) $\alpha(0) = 1$.

d) *Ist β ein Automorphismus von $(K, +, \cdot)$, so gilt $\alpha\beta = \beta\alpha$.*

Ist K ein Neokörper, so können b) *und* b') *ersetzt werden durch*

b'') *$x \to \alpha(x)\, x^{-1}$ ist eine eineindeutige Abbildung von K' auf $K - \{1\}$;*

dies ist eine Folgerung aus

d') *Ist β ein innerer Automorphismus von $(K', \cdot)$, so gilt $\alpha\beta = \beta\alpha$.*

Alle inneren Automorphismen β von $(K, \cdot)$ sind nämlich Automorphismen von $(K, +, \cdot)$. Ist umgekehrt eine Loop $(K, \cdot)$ gegeben, und gibt es ein nicht in K' gelegenes Element 0 und eine Abbildung α mit den Eigenschaften a), b) und c) für $K = K' \cup \{0\}$ und gilt d) für alle Abbildungen β mit $\beta(0) = 0$ von K auf sich, so daß die Einschränkung von β auf K' ein Automorphismus von $(K', \cdot)$ ist, so wird K durch die Festsetzung $x + y = y$, wenn $x = 0$ und $x + y = \alpha(y/x)\, x$ zu einer rechtsdistributiven Doppelloop $(K, +, \cdot)$, und wenn überdies b') gilt, zu einer distributiven; die Abbildungen β sind dann Automorphismen

von $(K, +, \cdot)$. Ist $(K', \cdot)$ eine Gruppe, und gelten a), b''), c) und d'), so ist $(K, +, \cdot)$ ein Neokörper (vgl. auch PAIGE [12], S. 41).

(4.2) *Die Addition ist genau dann kommutativ, wenn*

a) $\alpha(x/y)\,y = \alpha(y/x)\,x$ *für alle* $x \neq 0$ *und* $y \neq 0$

gilt. Ist $(K', \cdot)$ *assoziativ, dann ist die Addition genau dann kommutativ, wenn*

a') $\alpha(x)\,x^{-1} = \alpha(x)$ *für alle* $x \neq 0$

gilt. In diesem Fall wird die Bedingung (4.1) b) (bzw. b'')) *überflüssig.*

(4.3) *Die Addition ist genau dann linkskürzbar, d.h. sie genügt der Identität* $-x + (x + y) = y$ *für alle* $(x, y) \in K \times K$ *genau dann, wenn folgende Bedingung erfüllt ist:*

a) *Es gibt ein Element* -1 *in* K' *mit* $(-1)(-1) = 1$, *so daß die Abbildung* $\bar{\alpha}$ *mit* $\bar{\alpha}(x) = (-1)\,\alpha(x)$ *eine Involution ist.*

Ist $(K', \cdot)$ *assoziativ, so ist die Addition genau dann kürzbar,* d.h. *sie ist linkskürzbar und genügt der Identität* $(x + y) - y = y$ *für alle* $(x, y) \in K \times K$, *wenn* (4.2) a') *und* (4.3) a) *gilt.*

Beweis. (1.8), Satz von PICKERT.

(4.4) $(K, +)$ *ist genau dann potenzassoziativ, wenn* $f_{n,i}(1, \ldots, 1) = \alpha^{n-1}(1)$ *für alle* $n = 1, 2, \ldots, i \in I_n$ *gilt* [vgl. (1.9)]. *Dies ist gleichwertig mit der Gültigkeit aller Identitäten* $\alpha^n(1) = \alpha^k(1) + \alpha^l(1)$ *mit* $k + l = n + 1$, *was sich auch in der Form* $\alpha^n(1) = \alpha\left(\alpha^k(1)/\alpha^l(1)\right)\alpha^l(1)$ *mit* $k + l = n + 1$ *schreiben läßt.*

(4.5) *Nun sei* $(K, +, \cdot)$ *eine topologische distributive Doppelloop. Dann gilt für die Abbildung* α

a) α *ist ein Homöomorphismus.*

b) *Für alle* $(x_0, y_0) \in K \times K$ *existiert* $\lim (x/y)\,y$ *bei* $x \to x_0$, $y \to y_0$ *mit* $y \neq 0$, *und dieser Limes ist gleich* x_0, *wenn* $y_0 = 0$ *ist und gleich* $(x_0/y_0)\,y_0$ *sonst.*

c) α *ist mit allen (topologischen) Automorphismen von* $(K, +, \cdot)$ *vertauschbar.*

Ist umgekehrt eine topologische Loop $(K, \cdot)$ gegeben, und existiert ein topologischer Raum $K = K' \cup \{0\}$ und ein Homöomorphismus α von K, der (4.1) b), c) und (4.5) b) erfüllt, ist ferner die Einschränkung von α auf K' mit einem (topologischen) Automorphismus β von $(K', \cdot)$ vertauschbar, für den $\lim_{x \to 0} \beta(x) = 0$ gilt, so wird $(K, +, \cdot)$ eine topologische distributive Doppelloop durch die Festsetzung $y + x = x$, wenn $y = 0$ und $y + x = \alpha(x/y)\,y$, wenn $y \neq 0$; und β ist ein (topologischer) Automorphismus von $(K, +, \cdot)$. Gilt auch (4.1) b'), so ist $(K, +, \cdot)$ distributiv.

Wir wenden uns zunächst dem wiederum weniger übersichtlichen nicht zusammenhängenden Fall zu; dabei werden wir uns im wesentlichen auf Neokörper mit kommutativer Multiplikation beschränken. Ist nun $(K, +, \cdot)$ ein lokalkompakter total unzusammenhängender topologischer Neokörper mit kommutativer Multiplikation, so ist $(K', \cdot) = \langle \pi \rangle M$, und M ist eine offene Untergruppe, deren Typ wir in (3.3) gekennzeichnet haben. Jedes Element $x \in K'$ hat die Form $x = \pi^{w(x)} \cdot m(x)$ mit ganzem $w(x)$ und einem Element

$m(x)$ aus M. Die Abbildung $x \to w(x)$ ist ein Homomorphismus von $(K', \cdot)$ auf die additive Gruppe der ganzen Zahlen. Wir bezeichnen mit G die Menge $\{x;\ x = \mathbf{0}$ oder $x \in K', w(x) \geqq 0\}$. Die offenen und kompakten Teilräume $\pi^n G$, $n = 0, 1, \ldots$ bilden eine Basis der Umgebungen von $\mathbf{0}$. Für alle positiven n ist $(\pi^n G)\ G = G(\pi^n G) = \pi^n G$. Es gibt eine kleinste nichtnegative ganze Zahl l, so, daß $\pi^l + G \subseteqq G$, da G offen und kompakt ist. Wegen $G(\pi^l + G) = \pi^l G + G$ gilt dann sogar $\pi^l G + G = G$ und allgemeiner $\pi^{l+n} G + \pi^n G = \pi^n G$ für alle ganzen n. In derselben Weise folgt mit einer geeigneten kleinsten nichtnegativen ganzen Zahl r die Gleichung $\pi^n G + \pi^{n+r} G = \pi^n G$ für alle n. Es existiert demnach eine beschränkte, ganzwertige Funktion h einer ganzen Variablen, die außerhalb des Intervalls $[-r, l]$ verschwindet, so daß $w(x+y) = \min\big(w(x), w(y)\big) - h\big(w(x) - w(y)\big)$ gilt; insbesondere ist $w(x+y) \geqq \min\big(w(x), w(y)\big) - c$ mit einer festen ganzen Zahl c. Im Fall einer assoziativen Addition ist $c = 0$; dann ist nämlich G der Ring der ganzen Elemente und πG das Bewertungsideal zum Primelement π. Im allgemeinen Fall können wir jedoch nur feststellen, daß jede kompakte Unterloop von $(K, +)$ in genau einer Menge $\pi^n G$ mit maximalem n enthalten ist, und daß mit den für $(K, +)$ wie in (2.2) definierten Funktionen $(x_1, \ldots, x_m) \to g_i(x_1, \ldots, x_m)$ stets $g_{i+1}(G, \pi^r G, \ldots, \pi^r G) \subseteqq G$ gilt.

Ist nun $(K, +)$ potenzassoziativ, d.h. ist (4.4) erfüllt, so ist die von $\mathbf{1}$ erzeugte Unterloop von $(K, +)$ kompakt, denn sie ist in $\pi^{-r} G$ enthalten; ist nämlich $-\mathbf{1} + \mathbf{1} = \mathbf{0}$, so gilt $(-\mathbf{1})(-\mathbf{1}) = \mathbf{1}$, also $-\mathbf{1} \in M \subseteqq B$, was neben $\mathbf{1} \in B$ zu der behaupteten Relation ausreicht. Nach (1.2) ist $\mathbf{1}$ in einem (eventuell erst nach topologischer Abschließung) lokalkompakten Unterkörper H von K enthalten, der einen Primkörper R_0 der Charakteristik p enthält. H ist total unzusammenhängend und ist daher isomorph einer endlichen algebraischen Erweiterung eines vollständigen p-adischen Körpers oder isomorph einem Primkörper der Charakteristik $p > 0$ oder einem Potenzreihenkörper in einer Erzeugenden über einem solchen (vgl. PONTRJAGIN [15], S. 189). Es sei π' das bis auf Einheiten eindeutig bestimmte Primelement von H. Bezeichnet M' die Einheitengruppen von $(H', \cdot)$, so gilt $M' \subseteqq M$ und $H' = \langle \pi' \rangle M'$. Nun liegt π' in genau einer Klasse modulo M, d.h. es gibt eine natürliche Zahl e und ein $x \in M$, so daß $\pi' = \pi^e \cdot x$, und wir können annehmen, daß x nicht in M' liegt, denn π' ist nur modulo M' bestimmt. Da über die Divisionsverhältnisse in M/M' nichts bekannt ist, kann diese Aussage nicht verschärft werden. Wir stehen dabei vor denselben Schwierigkeiten wie in (3.17), wo wir in einem beliebigen total unzusammenhängenden Neokörper von dessen Zentrum zur vollen Multiplikationsgruppe aufgestiegen sind.

Ist nun x ein Element von K', dessen sämtliche nichtnegativen Potenzen in einer abelschen Untergruppe von $(K, +)$ liegen, so ist in $(K, +, \cdot)$ ein algebraisch homomorphes Bild H von $\overline{R}_0[X]'$, dem Ring aller Polynome $\sum_{n=-N}^{M} r_n X^n$ mit $r_n \in \overline{R}_0$ enthalten, wobei $\overline{R}_0$ die Hülle von R_0 in $(K, +, \cdot)$ bezeichnet, die also einem Primkörper der Charakteristik $p \neq 0$ oder einem vollständigen p-adischen Körper isomorph ist. Diese Behauptung ist ebenso zu

beweisen wie (1.19), wenn man berücksichtigt, daß H_0 abgeschlossen ist und R_0 in $\overline{R}_0$ dicht ist, so daß also auch $\overline{R}_0 H_0 = H_0$ ist. Nun ist aber H ein $\overline{R}_0$-Vektorraum und $\overline{H}$ ist lokalkompakt. Hat also $\overline{R}_0$ die Charakteristik 0, so ist notwendig H endlichdimensional über $\overline{R}_0$ und H ist abgeschlossen. Wird umgekehrt $R_0[X]'$ isomorph abgebildet, so hat R_0 die Charakteristik $p \neq 0$. Wir nehmen nun ohne die Allgemeinheit dabei einzuschränken an, daß x in $A_0 \cup B$ liegt (A_0, A_a und B wie in § 3), andernfalls würden wir zu x^{-1} übergehen können um jene Voraussetzung zu erzwingen. Nun liegen also auch alle Potenzen x^n in $A_0 \cup B$. Läge aber x in B, so läge jedes Polynom $p(x)$ in B, denn es gilt $p(x)^{p^n} = p(x^{p^n})$ und es wäre $H \subseteq B \cup \{0\}$: Dann müßte H aber diskret sein, da 0 in $(H, +)$ isoliert wäre; H wäre also auch abgeschlossen und kompakt, da in $B \cup \{0\}$ enthalten; schließlich wäre daher H endlich, was aber mit der Isomorphie zu $R_0[X]'$ unvereinbar ist. Damit haben wir gefunden, daß x in A_0 liegen muß. — Wegen der schon erwähnten Gleichung $p(x)^{p^n} = p(x^{p^n})$ ist ein Polynom $p(x)$ genau dann in A_0 enthalten, wenn $p(x)$ kein Absolutglied hat. Da jedes Polynom $p(x)$ in $A_0 \cup B$ liegt, gelten für das isomorphe Bild F von $R_0[X]$ die Beziehungen $F \subseteq A_0 \cup B$ und $xF \subseteq A_0 \cap F$. Die offenen Teilräume $x^n F$ bilden eine Umgebungsbasis des Nullelements in F. Betrachtet man nun auf $R_0[X]$ die durch die Exponentenbewertung $w(r_0 + r_1 X + \cdots + r_n X^n) = n$ gegebene Topologie, so ist der Isomorphismus von $R_0[X]'$ auf H sogar ein Homöomorphismus, da bei ihm die Umgebungsbasis $X^n \cdot R_0[X]$; $n = 1, \ldots$ der Null in $R_0[X]'$ auf die Umgebungsbasis $\{x^n F;$ $n = 1, \ldots\}$ der Null in F abgebildet wird. Die Vervollständigung von $R_0[X]'$ in der betrachteten Topologie ist der Körper $R_0\{X\}$ der nach unten abbrechenden Potenzreihen. Eine Cauchy-Folge in $R_0\{X\}$ geht in eine Cauchy-Folge in F über, die in $\overline{H}$ gegen einen Punkt von K konvergiert. Der Isomorphismus läßt sich also stetig fortsetzen zu einem Homomorphismus von $R_0\{X\}$ in $\overline{H}$. Da $R_0\{X\}$ ein Körper ist, hat jener den Kern $\{0\}$, und $R_0\{X\}$ wird durch ihn (topologisch) isomorph in und sogar auf $\overline{H}$ abgebildet.

Wir formulieren unsere Ergebnisse in Satz

(4.6) *$(K, +, \cdot)$ sei eine lokalkompakte totalunzusammenhängende topologische distributive Doppelloop.*

A) *Ist die Multiplikation eine kommutative Gruppe, dann gibt es einen Homomorphismus $x \to w(x)$ von $(K', \cdot)$ auf die additive Gruppe der ganzen Zahlen mit $w(x+y) \geqq \min(w(x), w(y)) - c$ mit einer festen ganzen Zahl c. Liegt $|w(x) - w(y)|$ oberhalb einer festen natürlichen Zahl n_0, so gilt sogar $w(x+y) \geqq \min(w(x), w(y))$.*

B) *Liegt ein Element $x \neq 0$ in einer Untergruppe von $(K', \cdot)$, so enthält K einen Unterkörper H, der entweder einem Primkörper der Charakteristik $p > 0$ oder einem p-adischen Körper isomorph ist. Ist ein Element x nicht in H enthalten, und liegen alle Potenzen von x in einer kommutativen Untergruppe von $(K, +)$, so enthält K einen Unterkörper $H(x)$, der im Falle der Charakteristik 0 stets eine endliche algebraische Erweiterung des Primkörpers ist und im Fall einer von 0 verschiedenen Charakteristik eine transzendente Erweiterung sein kann.*

Ist insbesondere $(K, +)$ diassoziativ, und gilt $x^2 = h \in H$, so ist $H(x)$ im erwähnten Sinne definiert

C) *Ist $(K, +)$ diassoziativ und ist die Charakteristik von K gleich 0, so ist $(K, +)$ kommutativ.*

Beweise. Um B) für einen beliebigen Neokörper mit nicht notwendig kommutativer Multiplikation einzusehen, braucht man sich nur auf das nach §3 existierende Zentrum einzuschränken, für welches die vorangegangenen Erläuterungen richtig bleiben. Der Teil C) folgt aus dem nachstehenden Hilfssatz, der sich als Gegenstück erweist zu einem Satz über alternative Fastvektorräume über dem Körper der reellen Zahlen aus [4], den wir in (4.10) noch einmal erwähnen.

(4.7) *$(G, +)$ sei ein diassoziativer topologischer Fastvektorraum über einem vollständigen p-adischen Körper K, dann ist $(G, +)$ kommutativ.*

Beweis. In der Identität $p^n \cdot (x + y) = x + y + \cdots + x + y = x + (p^n - 1)(x + y) + y$ lassen wir n gegen unendlich, also p^n gegen 0 streben. Dann bekommen wir $x - y - x + y = \mathbf{0}$, oder $x + z = z + x$ mit $z = -y$.

Im folgenden behandeln wir den zusammenhängenden Fall, und zwar zunächst die unterste Stufe, nämlich die reellen Neokörper, die als Zentrum in allen zusammenhängenden distributiven Doppelloops enthalten sind. Da die einem reellen Neokörper vermöge (4.1) und (4.5) eindeutig zugeordnete stetige streng monotone reelle Funktion α einer wohlvertrauten Klasse reeller Funktionen angehört, werden wir die reellen Neokörper mit aller wünschenswerten Anschaulichkeit überblicken können.

(4.8) *$x \to \alpha(x)$ sei eine stetige streng monotone Abbildung von $(R, +, \cdot)$, dem Körper der reellen Zahlen, auf sich mit $\alpha(0) = 1$. Dabei sei $x \to \frac{a(x)}{x}$ eine streng monotone Abbildung von R' auf $R - \{1\}$. Es existiere $\lim_{\substack{x \to x_0 \\ y \to 0}} \alpha\left(\frac{x}{y}\right) y$ und sei gleich x_0. Dann und nur dann definiert α auf dem Raum R einen reellen Neokörper $(R, *, \cdot)$ durch $y * x = \alpha\left(\frac{x}{y}\right) \cdot y$, wenn $y \neq 0$ und $y * x = x$, wenn $y = 0$.*

Beweis. (4.1) und (4.5).

Da jeder beliebige reelle Neokörper sich durch einen Homöomorphismus so auf den Körper der reellen Zahlen abgebildet werden kann, daß die Multiplikation isomorph mitgenommen wird, sind dadurch alle beliebigen reellen Neokörper erfaßt.

Die Funktion α ist fast überall differenzierbar und es gilt $\alpha'(x) \geqq 0$ und $x\alpha'(x) - \alpha(x) \leqq 0$ bzw. $\alpha'(x) \leqq 0$ und $x\alpha'(x) - \alpha(x) \geqq 0$. Der Graph von α hat die Asymptote $y = px + q$ mit reellen Zahlen $p \neq 0$ und q, denn $y = \alpha(x) x^{-1}$ hat wegen $\alpha(0) = 1$ und der geforderten Eineindeutigkeit eine waagrechte Asymptote. Eine Funktion mit den genannten Eigenschaften nennen wir zulässige Funktion ist z.B. $\alpha(x) = 1 + x + \mathfrak{Tg}\, x$. Sie liefert einen reellen Neokörper, dessen Addition nicht kommutativ ist und keinen Kürzungsregeln genügt. Im Fall endlicher Neokörper mit abelscher Multiplikation ist dies unmöglich (vgl. [12] oder [13], S. 91).

Ist die Addition kommutativ, so gilt nach (4.2) a') $\frac{\alpha(x)}{x} = \alpha\left(\frac{1}{x}\right)$. Die reelle Zahl -1 mit $(-1)^2 = 1$ ist das eindeutig bestimmte Inverse von 1 in $(R, *)$; daher muß α monoton steigend sein. Die Abbildung $x \to x^{-1}$ bildet das Äußere des Intervalls $(-1, 1)$ auf sein Inneres ab. Es genügt also, um eine zulässige Funktion anzugeben, eine eigentlich monotone Funktion g auf $[-1, 1]$ aufzuweisen mit $g(-1) = 0$, $g(0) = 1$ für welche auch $x \to \frac{1}{x} g(x)$ auf $[-1, 1] - \{0\}$ monoton ist, und dann

$$\alpha(x) = \begin{cases} g(x) & \text{für } x \in [-1, 1] \\ x\, g(x^{-1}) & \text{für } x \notin [-1, 1] \end{cases}$$

zu setzen.

Jeder Funktion auf $[-1, 1]$ mit den beschriebenen Eigenschaften ist also genau ein reeller Neokörper mit kommutativer Addition zugeordnet. Die Addition ist genau dann monassoziativ, wenn $g(x) = 1 + x$ für $x \geqq 0$ gilt (vgl. [5], § 3). Das in der angegebenen Arbeit des Verfassers erwähnte Beispiel *eines reellen Neokörpers mit kommutativer potenzassoziativer und nicht assoziativer Addition* ist durch

$$g(x) = \begin{cases} 1 + x & \text{für } x \in [0, 1] \\ \frac{1 + x^3}{1 + x^2} & \text{für } x \in [-1, 0) \end{cases}$$

gegeben. Die Linkskürzungsregel für die Addition gilt genau dann, wenn die Abbildung $x \to \bar{a}(x) = -\alpha(x)$ eine Involution ist. Dies ist aber genau dann der Fall, wenn der Graph von $x \to \alpha(x)$ bei der Spiegelung an der zweiten Winkelhalbierenden in sich übergeht.

Ein Beispiel eines reellen Neokörpers mit kommutativer und kürzbarer nichtassoziativer Addition gekennzeichnet durch

$$g(x) = \begin{cases} 2(x + 1) & \text{für } x \in [-1, -\frac{2}{3}] \\ \frac{1}{2} x + 1 & \text{für } x \in (-\frac{2}{3}, 1] \end{cases}$$

wurde von SALZMANN [16] angegeben.

Wir geben nun noch ein Beispiel *eines reellen Neokörpers, dessen Addition rechtskürzbar und monassoziativ, aber nicht assoziativ ist*, indem wir definieren

$$\alpha(x) = \begin{cases} 1 + x & \text{für } x \notin [-1, 0) \\ (1 - x^3)^{\frac{1}{3}} & \text{für } x \in [-1, 0). \end{cases}$$

Die Potenzassoziativität hat allerdings die Assoziativität der Addition zur Folge. Das Beispiel hat in der Theorie der abstrakten Loops insofern Interesse, als es zeigt, daß in einer Loop mit Rechtskürzungsregel wohl alle Elemente in einer Halbgruppe liegen können ohne daß auch nur eines in einer Gruppe liegt. Wir vermerken noch ein Beispiel *eines reellen Neokörpers mit nicht assoziativer Addition, in dem alle Verknüpfungen einmal (allerdings nicht stetig) differenzierbar sind*. Einsolches wird gegeben durch $\alpha(x) = x + \frac{1}{2}\left(1 + \frac{1}{1 + x^2}\right)$.

(4.9) *Es sei* $(K, +, \cdot)$ *eine zusammenhängende topologische distributive Doppelloop;* $\mathbf{1}$ *liege in einer additiven Gruppe.*

Dann hat K *die Charakteristik* 0.

Der Beweis folgt unmittelbar aus (2.5).

Das Zentrum von $(K, +, \cdot)$ ist ein komplexer Neokörper, der die Hülle $\overline{R}_0$ des Primkörpers R_0 enthält, die dem Körper der reellen Zahlen isomorph ist.

Nun sei x ein Element von K' das in einer Unterhalbgruppe von $(K', \cdot)$ liege und zusammen mit allen nichtnegativen Potenzen in einer abelschen Untergruppe von $(K, +)$. Dann ist x algebraisch über der Hülle $\overline{R}_0$ des Primkörpers. Diese Behauptung ist wieder ebenso zu beweisen wie (1.19), weil infolge der Abgeschlossenheit von H_0 auch $\overline{R}_0 \cdot H_0 = H_0$ gilt. Das Bild von $\overline{R}_0[X]$ bei dem Homomorphismus, der X in x abbildet ist ein in K enthaltener $\overline{R}_0$-Vektorraum und ist daher endlichdimensional, woraus die Behauptung folgt. Da $\overline{R}_0$ maximalgeordnet ist, gibt es darüber nur algebraische Erweiterungen vom Grad höchstens 2.

Ist insbesondere $(K, +)$ alternativ, und gibt es ein Element $x \in K$ mit $x^2 = -\mathbf{1}$, so ist x in einem Unterkörper enthalten, der zum Körper der komplexen Zahlen isomorph ist. Daraus folgt der Satz:

(4.10) *Sei* $(K, +, \cdot)$ *ein lokalkompakter zusammenhängender Neokörper;* $(K, +)$ *sei diassoziativ.* x *sei ein monassoziatives Element aus* $(K', \cdot)$ *mit der genauen Ordnung* 4. *Dann liegt* x *in einem Unterkörper* H *von* K, *der zum Körper der komplexen Zahlen isomorph ist. Ist ferner* T *eine eindimensionale Torusuntergruppe von* $(K', \cdot)$ *aus dem Kern, so liegt* T *in einem Unterkörper, der zum Körper der komplexen Zahlen isomorph ist.*

Beweis. $x^4 = \mathbf{1}$ hat $(x^2)^2 = \mathbf{1}$ zur Folge, was wegen $x^2 \neq \mathbf{1}$ gerade $x^2 = -\mathbf{1}$ nach sich zieht. Der erste Teil des Satzes folgt dann aus dem Vorhergehenden. O.B.d.A. sei $\dim K = 4$; falls $\dim K = 2$ ist, muß $K = H$ sein. Ist T_H die Torusuntergruppe von $(H', \cdot)$, so gilt $T_H \cap T \supseteq \langle x \rangle$ und alle eindimensionalen Torusuntergruppen von H, die durch $\langle x \rangle$ laufen, definieren denselben Unterkörper H. Die sämtlichen multiplikativen Torusuntergruppen des Kerns $(C', \cdot)$ von $(K', \cdot)$ überdecken die maximale kompakte Untergruppe von C', die zur einfach zusammenhängenden Überlagerungsgruppe der $SO(3)$ isomorph ist, schlicht mit Ausnahme der Punkte $\mathbf{1}$ und $-\mathbf{1}$; alle Torusuntergruppen, die T_H in $\langle x \rangle$ schneiden, sind daher gleich T. Also ist $T_H = T$ und die Behauptung folgt.

(4.11) *Ist* $(G, +)$ *ein diassoziativer* K*-Fastvektorraum über dem Körper* K *der reellen Zahlen, dann ist* $(G, +)$ *kommutativ.*

Beweis.

$$x + y = \frac{1}{n} \cdot (x + y + \cdots + x + y) = \frac{1}{n} \cdot x + \frac{n-1}{n} \cdot (y + x) + \frac{1}{n} \cdot y$$

liefert bei Übergang von n gegen unendlich die Behauptung.

Nun sei $(K, +\cdot)$ ein quaternaler Neokörper mit diassoziativer Addition. Wir denken uns K auf den Schiefkörper Q der Quaternionen durch h so

abgebildet, daß $(K', \cdot)$ zu $(Q', \cdot)$ isomorph ist und $h(0) = 0$ gilt. Die Zentren Z'_K und Z'_Q sind eindeutig bestimmt und gehen bei h ineinander über. Das Zentrum Z'_K hat die Form $L \times D$ mit einer zu $(R, +)$ isomorphen Einparametergruppe L, und einer Gruppe $D = \{1, -1\}$ aus zwei Elementen. Als stetige Automorphismen hat es die Abbildungen auf sich, die D festlassen und die L strecken, d.h. die den Streckungen $r \to c \cdot r$ von $(R, +)$ mit $c \neq 0$ entsprechen. Da $K' = L \times M$ ist (3.16), können wir durch Anwendung eines Automorphismus von K' den Homomorphismus h so normieren, daß $1_K + 1_K$ in $1_Q + 1_Q = 2$ übergeht. Es sei nun T_K eine eindimensionale Torusgruppe in $(K', \cdot)$. Dann ist $T_K \cap Z'_K = \{1, -1\}$ und $H_K = Z'_K T_K \cup \{0\}$ ist mit der in K erklärten Addition und Multiplikation ein komplexer Körper. Dasselbe gilt für $H_Q = h(Z'_K T_K) \cup \{0\}$ in Q. Es gibt einen Isomorphismus h' von $(H_Q, +, \cdot)$ auf $(H_K, +, \cdot)$. Die Abbildung $h \circ h'$ ist ein Automorphismus der Gruppe $h(Z'_K) \cdot h(T_K) = Z'_Q \cdot h(T_K)$, der Z'_Q und $h(T_K)$ als ganze festläßt. Die eindimensionale Torusgruppe $h(T_K)$ besitzt nur die identische Abbildung und die Abbildung $r \to -r \pmod 1$ (wenn man sie als additive Gruppe der reellen Zahlen modulo 1 auffaßt) als stetige Automorphismen. Da $h \circ h'$ das Element $2 \in Q \cap h(Z'_K)$ festläßt, bleiben alle Elemente von $h(Z'_K)$ fest. Bei der Wahl von h' haben wir die Möglichkeit, vorher den einzigen nichtidentischen stetigen Automorphismus von $(H_Q, +, \cdot)$ anzuwenden, der das Element i mit $i^2 = -1$ in $-i$ überführt. Dieser induziert auf der Torusgruppe $h(T_K)$ von $(H'_Q, \cdot)$ gerade den nichttrivialen Automorphismus. Aus diesem Grunde können wir annehmen, daß $h \circ h'$ auch $h(T_K)$ elementweis festläßt. Dann ist $h \circ h'$ aber die identische Abbildung, woraus folgt, daß h sogar ein Isomorphismus von $(H_K, +, \cdot)$ auf $(H_Q, +, \cdot)$ ist. Nach der angegebenen Normierung von h werden die zum komplexen Zahlkörper isomorphen Unterkörper von K isomorph auf diejenigen von Q abgebildet und alle komplexen Unterkörper von Q kommen dabei als Bilder vor. Durch $x * y = h\left(h^{-1}(x) + h^{-1}(y)\right)$ definieren wir eine alternative Addition auf Q, so daß $(Q, *, \cdot)$ zu $(K, +, \cdot)$ isomorph ist. Die Funktion $z \to r(z)$ sei bestimmt durch $r(z) = 1$ für $z = -1$ und durch $r(z) = (1 * z)^{-1}(1 + z)$ sonst. Dann ist zunächst $r(z) = 1$ für alle $z \in Z_Q$, wenn Z_Q das Zentrum von Q bezeichnet. Es sei H ein komplexer Unterkörper von Q, dann ist $(H, *, \cdot)$ identisch gleich $(H, +, \cdot)$. Denn für $(x, y) \in H \times H$ gilt ja $x * y = h\left(h^{-1}(x) + h^{-1}(y)\right) = h\left(h^{-1}(x + y)\right) = x + y$. Damit ist $r(z) = 1$ sogar für alle $z \in H$. Ist z_0 ein beliebiges Element von $Q - Z$, so geht durch die zweipunktige Menge $Z z_0 \cap M$ genau eine Torusgruppe. Nach (4.10) liegt also dann z_0 in einem komplexen Unterkörper von Q. Daher ist $r(z_0) = 1$. Nun ist $(Q, *, \cdot)$ als identisch zu $(Q, +, \cdot)$ erkannt. Demnach ist $(K, +, \cdot)$ isomorph zu $(Q, +, \cdot)$. Damit haben wir folgenden Satz bewiesen:

SATZ V

(4.12) *$(K, +, \cdot)$ sei eine lokalkompakte endlichdimensionale zusammenhängende topologische distributive Doppelloop mit diassoziativer Addition. Dann ist der Kern von K ein Korper, und zwar ein reeller, ein komplexer oder ein Quaternionenschiefkörper.*

Korollar:

(4.13) *Ein zusammenhängender lokalkompakter endlichdimensionaler Neokörper mit diassoziativer Addition ist einer der klassischen Pontrjaginschen Schiefkörper.*

Von der Diassoziativität der Addition wird man nicht abgehen können. Der Verfasser hat an anderer Stelle [5] ein Beispiel eines komplexen Neokörpers mit kommutativer und potenzassoziativer und kürzbarer aber nicht assoziativer Addition angegeben.

Für einen komplexen Neokörper K kann man eine zulässige Funktion $z \to \alpha(z)$ z.B. dadurch angeben, daß man für x eine in der ganzen Ebene schlichte Funktion nimmt, die 0 in 1 überführt. Eine solche hat die Form $\alpha(z) = 1 + pz$ mit $p \neq 0$. Das Element $-\frac{1}{p}$ ist also Linksinverses von 1 bezüglich der Addition. Setzen wir etwa $b = i$, so liegt die Menge $\alpha^n(1)$; $n = 0, 1, \ldots$ auf keiner multiplikativen Einparameteruntergruppe. Das hat zur Folge, daß der von 1 erzeugte Unterneokörper in K dicht ist. K ist also sein eigenes Zentrum und die in (3.15) erwähnte Einparametergruppe L ist in diesem Fall nicht eindeutig bestimmt [vgl. (3.19) A)].

Schlussbemerkungen

Die algebraische Theorie der Neokörper wurde zuerst von Paige [12] behandelt. Er untersucht vor allem die Frage, welche hinreichenden und notwendigen Bedingungen eine Gruppe erfüllen muß, damit ein Neokörper existiert, dessen multiplikative Gruppe zu der gegebenen Gruppe isomorph ist, und findet für den Fall endlicher und endlich erzeugter abelscher Gruppen ziemlich erschöpfende Antworten. Eine Fortsetzung dieser Arbeit sind die Untersuchungen über distributive Doppelloops, die von Hughes [6] angestellt wurden. In beiden Arbeiten interessiert besonders die Frage, wann ein Neokörper Koordinatenternärkörper einer affinen Ebene sein kann, eine Frage, die wir in dieser Arbeit wie alle mit der Geometrie zusammenhängenden Fragen nicht berührten. Topologische Neokörper traten zuerst bei Salzmann [16] auf, der feststellte, daß ein topologischer Neokörper, der zur reellen Geraden homöomorph ist, ein reeller Neokörper ist. Diesen Satz haben wir in Satz II etwas verschärft. Er gibt ferner ein Beispiel eines reellen Neokörpers mit kommutativer und kürzbarer, aber nicht assoziativer Multiplikation an [vgl. (4.8)].

Der Verfasser hat früher bewiesen, daß eine zweidimensionale topologische distributive Doppelloop mit potenzassoziativer Addition und Multiplikation ein komplexer Neokörper ist. Dieses Ergebnis ist in dem schärferen Satz III enthalten. In der erwähnten Arbeit war ferner gezeigt worden, daß eine lokalkompakte distributive Doppelloop mit Moufangscher Addition und kürzbarer Multiplikation, die außerdem keine beliebig kleinen additiven Untergruppen enthält entweder einer der klassischen zusammenhängenden Schiefkörper oder eine Cayleyalgebra ist. Dieses Ergebnis wurde unter entscheidender Verwendung der Moufang-Eigenschaft der Addition abgeleitet, während Satz IV ohne

Bezug auf die Addition bewiesen wurde. Der Satz V ist eine Ergänzung zu jenem Satz. Die Voraussetzungen sind abgeschwächt insofern, als bei der Addition nur die Diassoziativität vorausgesetzt wird, aber auch verschärft, da die Assoziativität der Multiplikation gefordert wird.

Die Beweise von (2.2) und (3.1) sind weitgehend Verallgemeinerungen und Spielarten von den Methoden, die PONTRJAGIN und KOWALSKY zur Untersuchung der Struktur lokalkompakter topologischer Körper verwenden. (2.2) wurde rein looptheoretisch formuliert und hatte so als interessante Konsequenz den Satz (2.5), der allerdings außerhalb des von uns hier dargestellten Zusammenhangs steht. Der Begriff des Fastvektorraums erweist sich als recht tragkräftiges Werkzeug zur Klärung der Einbettung des Zentrums in einer distributiven Doppelloop.

Die zur Untersuchung der Multiplikationsgruppe verwendeten Hilfsmittel aus der Theorie der lokalkompakten Gruppen liegen als solche nicht an der Oberfläche.

Literatur

[1] BRUCK, R. H.: Contribution to the Theory of Loops. Trans Amer. Math. Soc. **60**, 245–354 (1946). — [2] CHEVALLEY, C.: Theory of Liegroups. Princeton Univ. Press 1946. — [3] HOFMANN, K. H.: Topologische Loops. Math. Z. **70**, 13–37 (1958). — [4] HOFMANN, K. H.: Topologische Loops mit schwachen Assoziativitätsforderungen. Math. Z. **70**, 125–155 (1958). — [5] HOFMANN, K. H.: Topologische Doppelloops. Math. Z. **70**, 213–230 (1958). — [6] HUGHES, D. R.: Planar Division Neorings. Trans. Amer. Math. Soc. **80**, 502–527 (1955). — [7] KAPLANSKY, I.: Infinite Abelian Groups. Univ. of Mich. Publ. in Math. Nr. 2, Univ. of Michigan Press 1954. — [8] KOWALSKY, H. J.: Zur topologischen Kennzeichnung von Körpern. Math. Nachr. **9**, 261–268 (1953). — [9] MONTGOMERY, D., and L. ZIPPIN: Topological Transformation Groups. Intersc. Tracts in pure and appl. Math. **1956**. — [10] MOSTERT, G. D., and A. L. SHIELDS: On the Structure of Semigroupson a compact Manifold with Boundary. Ann. of Math. **65**, 117–143 (1957). — [11] MOSTERT, G.D.: On a compact Liegroup acting on a manifold. Ann. of Math. **65**, 447–455 (1957). — [12] PAIGE, G.: Neofields. Duke Math. J. **16**, 39–60 (1949). — [13] PICKERT, G.: Projektive Ebenen. Berlin: Springer, Gelbe Reihe LXXX, 1955. — [14] PICKERT, G.: Ebenen über Neokörpern. Wiss. Z. d. Fr.-Schiller-Universität Jena 1955/56. Math.-natw. Reihe, H. 1/2, 131–135. — [15] PONTRJAGIN, L. S.: Topologische Gruppen I. Leipzig: Teubner 1957. — [16] SALZMANN, H.: Projektive Ebenen. Math. Z. **67**, 436–466 (1957). — [17] WEIL, A.: L'Intégration dans les groupes topologiques. Act. Sc. Ind. **869–1145** (1939–1951).

Tübingen, Mathematisches Institut der Universität

(Eingegangen am 8. Juli 1958)

FOSTER, A. L.
Math. Zeitschr. 71, 69–82 (1959)

An Existence Theorem for Functionally Complete Universal Algebras

By
ALFRED L. FOSTER

1. Introduction. Let $Sp = (n_1, n_2, \ldots)$ be a given species and $\mathfrak{A} = (A, o_1, o_2, \ldots)$ a (universal) algebra of species Sp; here $A = \{\ldots, \zeta, \ldots\}$ denotes the class of elements of $\mathfrak{A}$ and $o_1, o_2, \ldots$ the primitive operations, where o_i is n_i-ary $(i = 1, 2, \ldots)$,

$$o_i(\zeta_1, \zeta_2, \ldots, \zeta_{n_i}) \in A \qquad (\zeta_1, \zeta_2, \ldots, \zeta_{n_i} \in A).$$

$\mathfrak{A}$ is called *functionally complete* (= f.c.) if for each positive integer k each mapping $f(\zeta_1, \zeta_2, \ldots, \zeta_k)$ of the set $A^{(k)}$ $(= k^{th}$ Cartesian power of $A)$ into the set A may be expressed as an $\mathfrak{A}$-*function*, i.e., as some primitive composition of k indeterminates $\zeta_1, \zeta_2, \ldots, \zeta_k$ and a (possibly empty) set of constants (= fixed elements) of A. If each mapping $f(\zeta_1, \zeta_2, \ldots, \zeta_k)$ may be expressed as a *strict* $\mathfrak{A}$-function — i.e., as an $\mathfrak{A}$-function in which no constants occur — the algebra $\mathfrak{A}$ is called functionally *strictly complete* (= f.s.c.). An algebra $\mathfrak{P} = (P, o_1, o_2, \ldots)$ which is f.s.c. and not isomorphic with the one-element algebra, 1, is called *primal* (of species Sp). A primal $\mathfrak{P}$ is necessarily of finite order (P = finite class); furthermore the n_i of its Sp cannot all be 1 (see [1], [2]).

In the present communication we establish the existence of a comprehensive class of primal algebras of species (2,1), a class which, to the best of the author's knowledge, subsumes all heretofore known instances of this species (see [1], [2], also the conclusion of this section). In the interest of a broader orientation we shall first briefly recall a recent field of application of primal algebras, of arbitrary species.

Corresponding to a given primal algebra, $\mathfrak{P}$, a "$\mathfrak{P}$-algebra" is defined as: any algebra (of the same species as $\mathfrak{P}$) which satisfies all strict identities satisfied by $\mathfrak{P}$. From [1] to [4] (and additional references found in [4]) a single illustration will point up the fundamental role played by primal algebras:

THEOREM 1. *Let $\mathfrak{P}$ be an arbitrary primal algebra. Then:*

(i) The class of all $\mathfrak{P}$-algebras is coextensive (up to isomorphisms) with the class of all subdirect powers of $\mathfrak{P}$ — or, equivalently, with the class of all subalgebras of direct powers of $\mathfrak{P}$.

(ii) In a $\mathfrak{P}$-algebra, $\mathfrak{A}$, each $\mathfrak{A}$-function possesses a normal expansion over $\mathfrak{P}$.

(iii) Each $\mathfrak{P}$-algebra possesses an n-ality theory, where n = order of $\mathfrak{P}$.

Theorem 1, for suitable specializations of the "kernel" $\mathfrak{P}$, subsumes familiar results of such disciplines as (a) Boolean algebras, (b) p-rings (MCCOY and MONTGOMERY [5], FOSTER [1], [2]), (c) Post algebras (ROSENBLOOM [6], WADE [7], FOSTER [1], [2]) and others (see [1]). Here the "Boolean kernel"

is $(F_2, \times, ^\frown)$ and more generally the "p-ring kernel" is $(F_p, \times, ^\frown)$, where $\times$ is the product and $\zeta^\frown = 1 + \zeta$ in the prime field $(F_p, \times, +)$; again the "Post-kernel" (of order n) is the algebra $(P_n, \times, ^\frown)$ where

$$P_n = \{0, \alpha_{n-2}, \alpha_{n-3}, \cdots, \alpha_2, \alpha_1, 1\}$$
$$\zeta \times \eta = \text{def} = \min(\zeta, \eta), \quad (\text{where } 0 < \alpha_{n-2} < \cdots < \alpha_1 < 1)$$
$$\zeta^\frown = \text{def: } 0^\frown = 1,\ 1^\frown = \alpha_1,\ \alpha_1^\frown = \alpha_2, \ldots, (\alpha_{n-2})^\frown = 0.$$

Each of the above kernels is known to be a primal algebra; additional examples, of arbitrary order n, are exhibited in [1], [2].

The profusion of primal algebras — subsuming all of the above — whose existence we here establish (§ 2), substantially broadens the field of application of theorem 1 and of additional such results given in [2], [3].

2. Binary algebras, frames; principal theorem. By a *binary algebra* $(B, \times)$ we understand a universal algebra of species (2) which possesses a null and an identity:

$$0, 1 \in B, \quad 0 \neq 1$$
$$0 \times \zeta = \zeta \times 0 = 0; \quad 1 \times \zeta = \zeta \times 1 = \zeta \qquad (\zeta \in B)$$

In general $\zeta \times \eta$ (also written simply $\zeta\eta$) is neither associative nor commutative.

An algebra $(B, \times, ^\frown)$ of species (2,1) is called a *cyclic frame* — hereafter simply called a *frame* — if (i) $(B, \times)$ is a binary algebra, (ii) $\zeta^\frown$ is a cyclic permutation of the set B and where (iii) $0^\frown = 1$. The inverse of the permutation $^\frown$ is denoted by $^\smile$; if there is a fixed n such that $\zeta^{\frown n} = \zeta$ for all $\zeta \in B$, as for instance when B is finite (= finite frame), then of course $\zeta^\smile = \zeta^{\frown n-1}$. (Here $\zeta^{\frown r}$ denotes the r^{th} iteration of $\zeta^\frown$, with $\zeta^{\frown 0} = \zeta$, $\zeta^{\frown 1} = \zeta^\frown$, $\zeta^{\frown 2} = (\zeta^\frown)^\frown$, etc.).

A finite frame $(B, \times, ^\frown)$ which is functionally strictly complete (and consequently a primal algebra, since $0 \neq 1$) is called a *primal frame*. It is noted that the several classes of kernels referred to following theorem 1 are all primal frames. How extensive is this class of primal algebras? We answer this by proving

THEOREM 2 (Principal theorem on primal frames).

(1) *Let $(B, \times)$ be a finite binary algebra, not isomorphic with the 4-element algebra $(H_4, \times)$ below but otherwise entirely arbitrary. Then there exists at least one cyclic permutation, $^\frown$, of the class B such that $(B, \times, ^\frown)$ is a primal frame.*

(2) *Let $(H_4, \times)$ denote the binary algebra given by*

$\times$	0	1	α	β
0	0	0	0	0
1	0	1	α	β
α	0	α	α	0
β	0	β	0	β

There exists no permutation, $^\frown$, of H_4 such that $(H_4, \times, ^\frown)$ is a primal frame.

The proof of theorem 2 requires some preparation, and is given via a number of lemmas. The proof is finally completed in § 7.

3. The characteristic functions. Let $\mathfrak{B}=(B, \times, \frown)$ be a (not necessarily finite) frame, and let $\beta \in B$. The functions $\delta_\beta(\zeta)$, $\Delta_\beta(\zeta)$ are called the β^{th} *characteristic* functions of the first and second kinds, respectively; here

$$\delta_\beta(\zeta) = \text{def} = \begin{cases} 1 \text{ for } \zeta = \beta \\ 0 \text{ for } \zeta \neq \beta; \end{cases} \qquad \Delta_\beta(\zeta) = \text{def} = \begin{cases} 0 \text{ for } \zeta = \beta \\ 1 \text{ for } \zeta \neq \beta. \end{cases}$$

For δ_0, Δ_0 (called the principal characteristic functions) we write simply $\delta_0(\zeta) = \delta(\zeta)$, $\Delta_0(\zeta) = \Delta(\zeta)$. In general, for a given frame $\mathfrak{B}$, none of the characteristic functions will be expressible as a $\mathfrak{B}$-function, much less as a strict $\mathfrak{B}$-function (i.e., some composition, via the primitive operations $\times, \frown$ of $\mathfrak{B}$, of an indeterminate ζ over B). We require the following result, taken from [*1*].

THEOREM 3. *Let $\mathfrak{B}=(B, \times, \frown)$ be a finite frame. Then each of the following implies the others, and constitutes a necessary and sufficient condition for $\mathfrak{B}$ to be functionally strictly complete (and therefore a primal frame):*

(1) *$\delta(\zeta)$ is expressible as a (not necessarily strict) $\mathfrak{B}$-function.*

(1′) *$\Delta(\zeta)$ is expressible as a (not necessarily strict) $\mathfrak{B}$-function.*

(2) *for at least one $\beta \in B$, $\delta_\beta(\zeta)$ is expressible as a $\mathfrak{B}$-function.*

(2′) *for at least one $\beta \in B$, $\Delta_\beta(\zeta)$ is expressible as a $\mathfrak{B}$-function.*

4. The rank and type of a binary algebra, etc. We introduce several closely related notions which will prove useful in the proof of theorem 2. Let $(B, \times)$ be a binary algebra. A *proper* element of B is one $\neq 0, \neq 1$. A *proper product* (abbreviated prpr) of B is a product of one or more proper elements; since $\times$ is generally not associative, nor commutative, both the association and the ordering must of course be specified in a product. The *length* of a product is the number of distinct proper elements of B which occur in it. Thus, for example, with α, β distinct and proper, $(\alpha\beta)\{(\beta^2\alpha)(\alpha^2\alpha)\}$ is a prpr of length 2.

A binary algebra $(B, \times)$ is said to be of *rank r* if there exists at least one non-vanishing (i.e. $\neq 0$) product of length r, while every product of length $>r$ vanishes. Thus, for instance, the algebra $(H_4, \times)$ of theorem 2 is of rank 1. Again, for example, $(K_4, \times)$ is of rank 2, if $K_4 = \{0, 1, \alpha, \beta\}$ and $\alpha\beta = \beta\alpha = 0$, $\alpha^2 = 1$, $\beta^2 = \beta$, since $(\alpha^2)(\beta^2) \neq 0$.

Let $\mathfrak{B}=(B, \times)$ be a finite binary algebra of n elements (=order n). From the definition we immediately have:

$$(4.1) \qquad \begin{cases} \text{rank } \mathfrak{B} = 0 & \text{if } n = 2 \\ 1 \leqq \text{rank } \mathfrak{B} \leqq n-2 & \text{if } n > 2. \end{cases}$$

Corresponding to each rank r we distinguish two *types*, *i* (= external-type) and *ii* (= internal-type). A finite binary algebra $(B, \times)$ of rank r, is defined to be

(i) of *type i* if there exists a prpr $\mathscr{P}(\alpha_1, \alpha_2, \ldots, \alpha_r)$ of length r such that

(4.2) $\quad \mathscr{P}(\alpha_1, \alpha_2, \ldots, \alpha_r) = \mu, \quad$ where $\quad \mu \neq 0, \ \neq \alpha_i \quad$ (for $i = 1, 2, \ldots, r$).

(ii) of *type ii* if it is not of type i. Otherwise stated, $(B, \times)$, of rank r, is said to be of type *ii* if each prpr $\mathscr{P}(\alpha_1, \ldots, \alpha_r)$ of length r either vanishes or is equal to some one of the "factors" α_i appearing in $\mathscr{P}$,

(4.3) $\quad \mathscr{P}(\alpha_1, \ldots, \alpha_r) = \mu$ implies $\mu = 0$ or $\mu =$ some $\alpha_i \quad$ (for $i = 1, 2, \ldots, r$).

5. Order < 4. In this and the following sections the proof of theorem 2 is broken up into an exhaustive class of special cases.

LEMMA 5. *Let $(B, \times)$ be a binary algebra of order 2 or 3. Then the conclusion of theorem 2 holds.*

Proof. There is only one binary algebra of order 2, $(B, \times)$ with $B = \{0, 1\}$, and of course only one cyclic ordering of B, $0^\frown = 1$, $1^\frown = 0$. That this $(B, \times, ^\frown)$ is indeed a primal frame follows from theorem 3 since clearly

$$\delta(\zeta) = \zeta^\frown.$$

Again, there are precisely three binary algebras of order 3, $(B, \times)$: $B = \{0, 1, \alpha\}$ with the three $\times$'s below, corresponding to the three values of μ,

$\times$	0	1	α
0	0	0	0
1	0	1	α
α	0	α	μ

$(\mu = 0, 1 \text{ or } \alpha)$.

In each of these three cases define $\zeta^\frown$ by the ordering,

$$\zeta^\frown = \text{def: } 0, 1, \alpha \qquad (\text{i.e., } 0^\frown = 1, \ 1^\frown = \alpha, \ \alpha^\frown = 0).$$

We then readily verify that

$$\Delta_1(\zeta) = (\alpha\zeta)^\frown = \{(\zeta\zeta^\frown\zeta^{\frown 2})^{\frown 2} \times \zeta\}^\frown \qquad \text{for } \mu(=\alpha^2) = 0$$

$$\Delta(\zeta) = \zeta^2 \qquad \text{for } \mu = 1$$

$$\Delta_1(\zeta) = (\zeta\zeta^\frown)^\frown \qquad \text{for } \mu = \alpha.$$

A reference to theorem 3 then completes the proof of lemma 5.

6. Rank $\geqq 2$, type *i*.

LEMMA 6. *Let $\mathfrak{B} = (B, \times)$ be a finite binary algebra of rank $r \geqq 2$ and of type i. Then the conclusion of theorem 2 holds.*

Proof. Because of lemma 5 we need (and shall) consider only the case where the

(6.1) $\qquad$ order, n, of $\mathfrak{B} \geqq 4$.

Let

(6.2) $\qquad B = \{0, 1, \beta_1, \beta_2, \ldots, \beta_{n-2}\}$.

Under the hypotheses obtaining there exists a prpr, $\mathscr{P}$, of length r, such that

$$\mathscr{P}(\alpha_1, \alpha_2, \ldots, \alpha_r) = \mu, \quad \mu \neq 0, \quad \mu \neq \alpha_i \quad (i = 1, 2, \ldots, r). \tag{6.3}$$

Case 1, $\mu = 1$.

In this case it follows that r has its maximal possible value

$$r = n - 2. \tag{6.4}$$

For otherwise we should have some $\alpha_{r+1} \in B$ where $\alpha_{r+1} \neq 0, \neq 1, \neq \alpha_i$ $(i = 1, 2, \ldots, r)$, from which follows

$$\mathscr{P}(\alpha_1, \alpha_2, \ldots, \alpha_r) \times \alpha_{r+1} = \alpha_{r+1}. \tag{6.5}$$

Since the left side of (6.5) is a prpr of length $r+1$ it must vanish, and we are left with the contradiction: $0 = \alpha_{r+1}$. Hence (6.4) is established.

Let now the cyclic permutation $\zeta^\frown$ be defined by the ordering

$$\zeta^\frown = \text{def: } 0, 1, \alpha_1, \alpha_2, \ldots, \alpha_{n-2} \quad (\text{i.e., } 0^\frown = 1,\ 1^\frown = \alpha_1, \ldots, (\alpha_{n-2})^\frown = 0) \tag{6.5.1}$$

and let

$$\mathscr{P}'(\zeta) = \text{def} = \zeta \times \mathscr{P}(\zeta^\frown, \zeta^{\frown 2}, \ldots, \zeta^{\frown n-2}). \tag{6.6}$$

We show that the characteristic function $\delta_1(\zeta)$ is here given by

$$\mathscr{P}'(\zeta) = \delta_1(\zeta). \tag{6.7}$$

This is seen since if ζ is assigned any value from B the left of (6.6) is a product of $n-1$ distinct $\in B$; for $\zeta = 1$, $\mathscr{P}'(1) = \mu = 1$, while for $\zeta \neq 1$ one of these $n-1$ factors is 0 and hence $\mathscr{P}'(\zeta) = 0$ $(\zeta \neq 1)$. Thus (6.7) is verified and case 1 is proved, by reference to theorem 3.

Case 2, $\mu \neq 1$.

Since

$$\mathscr{P}(\alpha_1, \ldots, \alpha_r) = \mu \qquad (\mu \neq 0, \neq 1, \neq \alpha_i \quad (i = 1, 2, \ldots, r)) \tag{6.8}$$

if follows that

$$r < n - 2. \tag{6.9}$$

Let now $\zeta^\frown$ be defined by the ordering

$$\begin{cases} \zeta^\frown = \text{def: } 0, 1, \alpha_1, \alpha_2, \ldots, \alpha_r, \alpha_{r+1}, \ldots, \alpha_{n-3}, \mu \\ \qquad (\text{i.e., } 0^\frown = 1,\ 1^\frown = \alpha_1, \ldots, (\alpha_{n-3})^\frown = \mu,\ \mu^\frown = 0), \end{cases} \tag{6.10}$$

where the (possibly empty set of) elements $\alpha_{r+1}, \ldots, \alpha_{n-3}$ are ordered in an arbitrary but fixed way. Let

$$\mathscr{P}'(\zeta) = \text{def} = \zeta \times \mathscr{P}(\zeta^\frown, \zeta^{\frown 2}, \ldots, \zeta^{\frown r}). \tag{6.11}$$

We shall verify that

$$\mathscr{P}'(\zeta) = \begin{cases} \mu & \text{if } \zeta = 1 \\ 0 & \text{if } \zeta \neq 1. \end{cases} \tag{6.12}$$

That $\mathscr{P}'(1) = \mu$ follows from (6.8), (6.10) and (6.11). Further, if for a given value of ζ all of the $r+1$ different elements $\zeta, \zeta^\frown, \zeta^{\frown 2}, \ldots, \zeta^{\frown r}$ are proper

(i.e., $\neq 0, \neq 1$), then for such ζ, $\mathscr{P}'(\zeta) = 0$, since by (6.11) we then have a prpr of length $r+1$. Obviously $\mathscr{P}'(\zeta)$ vanishes if any of the quantities $\zeta, \zeta^{\frown}, \zeta^{\frown_2}, \ldots, \zeta^{\frown_r}$ is 0. There remains only the case where none of the above quantities is 0 and one of them is 1; but this implies $\zeta = 1$, and $\mathscr{P}'(1) = \mu$ as already observed. Thus (6.12) is proved. Referring now to (6.10) we have

$$(\mathscr{P}'(\zeta))^{\frown} = \Delta_1(\zeta) \tag{6.13}$$

whence case 2 is proved by use of theorem 3. This completes the proof of lemma 6.

7. Rank $\geqq 2$, type *ii*.

LEMMA 7. *Let $\mathfrak{B} = (B, \times)$ be a finite binary algebra of rank $r \geqq 2$ and of type ii. Then the conclusion of theorem 2 holds.*

Proof. As earlier, because of lemma 5 we need only consider the case where the

$$\text{order, } n, \text{ of } \mathfrak{B} \geqq 4. \tag{7.1}$$

With an appropriate lettering of the elements we now have a prpr $\mathscr{P}$ such that

$$\mathscr{P}(\alpha_1, \alpha_2, \ldots, \alpha_r) = \alpha_1. \tag{7.2}$$

Case 1, $r = n-2$.

In this case define $\zeta^{\frown}$ by the ordering

$$\zeta^{\frown} = \text{def: } 0, 1, \alpha_{n-2}, \alpha_{n-3}, \ldots, \alpha_3, \alpha_2, \alpha_1. \tag{7.3}$$

Now let

$$\mathscr{P}'(\zeta) = \text{def} = \mathscr{P}(\zeta^{\frown_1}, \zeta^{\frown_2}, \ldots, \zeta^{\frown_{n-2}}) \times \zeta^{\frown_{n-1}}. \tag{7.4}$$

One readily verifies that $\mathscr{P}'(0) = \alpha_1$, $\mathscr{P}'(\zeta) = 0$ $(\zeta \neq 0)$. Therefore, using (7.3), we have

$$(\mathscr{P}'(\zeta))^{\frown} = \Delta(\zeta) \tag{7.5}$$

and case 1 is proved by theorem 3.

Case 2, $r = n-3$.

From (7.1) we now have

$$\mathscr{P}(\alpha_1, \ldots, \alpha_{n-3}) = \alpha_1. \tag{7.6}$$

Again define $\zeta^{\frown}$ by the ordering (7.3), but now set

$$\mathscr{P}'(\zeta) = \text{def} = \mathscr{P}(\zeta^{\frown_1}, \zeta^{\frown_2}, \ldots, \zeta^{\frown_{n-2}}) \times \zeta^{\frown_{n-1}}. \tag{7.7}$$

We have

$$\begin{cases} \mathscr{P}'(0) = \alpha_1 \\ \mathscr{P}'(\alpha_2) = \mathscr{P}(\alpha_3, \alpha_4, \ldots, \alpha_{n-2}, 1) \times \alpha_1 \\ \mathscr{P}'(\zeta) = 0 \ (\zeta \neq 0, \neq \alpha_2), \text{ for simple reasons}^1). \end{cases} \tag{7.8}$$

[1]) Here and in the sequel, "for simple reasons" signifies that: either one factor is 0, or that we have a prpr of length greater than the rank r.

From (7.6), (7.8) we have

$$\mathscr{P}'(\alpha_2) = \mathscr{P}(\alpha_3, \alpha_4, \ldots, \alpha_{n-2}, 1) \times \mathscr{P}(\alpha_1, \alpha_2, \ldots, \alpha_{n-2}) \tag{7.9}$$

from which follows that $\mathscr{P}(\alpha_2) = 0$, since the right of (7.9) is a product of length $n-2 > r$. Therefore

$$\mathscr{P}'(\zeta) = \begin{cases} \alpha_1 & \text{for } \zeta = 0 \\ 0 & \text{for } \zeta \neq 0. \end{cases} \tag{7.10}$$

Using (7.3),

$$(\mathscr{P}'(\zeta))^{\frown} = \Delta(\zeta) \tag{7.11}$$

and case 2 is proved, by theorem 3.

Case 3, $r = n-k-2$, $k \geqq 2$.

The elements of B may be written

$$B = \{0, 1, \alpha_1, \alpha_2, \ldots, \alpha_r, \beta_1, \beta_2, \ldots, \beta_k\}, \quad (r \geqq 2,\ k \geqq 2). \tag{7.12}$$

We first split this case 3 into two subcases, 3*a* and 3b, each of which we again separate into several subcases $3a_1, \ldots, 3a_5$; $3b_1, 3b_2$.

Case 3a: there exists a $\beta (= \beta_1, \beta_2, \ldots, \beta_k)$ and an α $(= \alpha_2, \alpha_3, \ldots, \alpha_r)$ such that

$$\beta\alpha = 0. \tag{7.13}$$

Case $3a_1$, $r = k$.

Arrange the notation so that

$$\beta_2\alpha_r = 0. \tag{7.14}$$

Define $\zeta^{\frown}$ by the ordering

$$\zeta^{\frown} = \text{def: } 0, 1, \beta_k, \beta_{k-1}, \ldots, \beta_2, \beta_1, \alpha_r, \alpha_{r-1}, \ldots, \alpha_2, \alpha_1. \tag{7.15}$$

Let

$$\mathscr{P}'(\zeta) = \text{def} = \mathscr{P}(\zeta^{\smile_1} \times \zeta^{\smile_{n-1}}, \zeta^{\smile_2}, \zeta^{\smile_3}, \ldots, \zeta^{\smile_r}). \tag{7.16}$$

From (7.2), (7.15), (7.16) we have

$$\begin{cases} \mathscr{P}'(0) = \alpha_1 \\ \mathscr{P}'(\beta_1) = \mathscr{P}(\beta_2 \times \alpha_r, \beta_3, \ldots, 1) = 0 \quad \text{(by (7.14))} \\ \mathscr{P}'(\zeta) = 0 \ (\zeta \neq 0, \neq \beta_1), \text{ (for simple reasons).} \end{cases} \tag{7.17}$$

Hence from (7.15)

$$(\mathscr{P}'(\zeta))^{\frown} = \Delta(\zeta), \tag{7.18}$$

a d case $3a_1$ is proved by theorem 3.

Case $3a_2$, $k < r$.

range the notation so that

$$\beta_k\alpha_k = 0. \tag{7.19}$$

Define $\zeta^\frown$ by the ordering (7.15), but now let

$$\mathscr{P}'(\zeta) = \text{def} = \mathscr{P}(\zeta^{\smile_1}, \zeta^{\smile_2}, \ldots, \zeta^{\smile_{r-1}} \times \zeta^{\smile_{n-1}}, \zeta^{\smile_r}). \tag{7.20}$$

From (7.2), (7.15), (7.20) we have

$$\begin{cases} \mathscr{P}'(0) = \alpha_1 \\ \mathscr{P}'(\alpha_{k+1}) = \mathscr{P}(\alpha_{k+2}, \alpha_{k+3}, \ldots, \beta_1\alpha_k, 1) = 0 \ \text{(by (7.19))} \\ \mathscr{P}'(\zeta) = 0 \ (\zeta \neq 0, \neq \alpha_{k+1}), \text{ (for simple reasons).} \end{cases} \tag{7.21}$$

Hence from (7.15),

$$(\mathscr{P}'(\zeta))^\frown = \Delta(\zeta) \tag{7.22}$$

and case $3a_2$ is proved by theorem 3.

Case $3a_3$, $2<r<k$.

Arrange the notation so

$$\beta_{k-1} \times \alpha_{r-1} = 0. \tag{7.23}$$

Define $\zeta^\frown$ by the ordering

$$\zeta^\frown = \text{def: } 0, 1, \alpha_r, \alpha_{r-1}, \ldots, \alpha_3, \alpha_2, \beta_k, \ldots, \beta_3, \beta_2, \beta_1, \alpha_1. \tag{7.24}$$

Let

$$\mathscr{P}'(\zeta) = \text{def} = \mathscr{P}(\zeta^{\smile_1}, \zeta^{\smile_{k+2}} \times \zeta^\frown, \zeta^{\smile_{k+3}}, \ldots, \zeta^{\smile_{k+r}}). \tag{7.25}$$

From (7.2), (7.24), (7.25) we have

$$\begin{cases} \mathscr{P}'(0) = \mathscr{P}(\alpha_1, \alpha_2 \times 1, \alpha_3, \ldots, \alpha_r) = \alpha_1 \\ \mathscr{P}'(\alpha_r) = \mathscr{P}(1, \beta_{k-1} \times \alpha_{r-1}, \ldots) = 0 \quad \text{(by (7.23))} \\ \mathscr{P}'(\zeta) = 0 \ (\zeta \neq 0, \neq \alpha_r), \text{ (for simple reasons).} \end{cases} \tag{7.26}$$

Hence from (7.24),

$$(\mathscr{P}'(\zeta))^\frown = \Delta(\zeta) \tag{7.27}$$

and case $3a_3$ is proved (theorem 3).

Case $3a_4$, $r=2$, $k=3$.

Here

$$B = \{0, 1, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3\}.$$

Arrange the notation so that

$$\beta_1\alpha_2 = 0. \tag{7.28}$$

We further observe that

$$\beta_2\alpha_1 = 0 \tag{7.29}$$

since, using (7.2),

$$\beta_2\alpha_1 = \beta_2 \times \mathscr{P}(\alpha_1, \alpha_2) = 0 \quad \text{(prpr of length 3)}.$$

Define $\zeta^\frown$ by the ordering

$$\zeta^\frown = \text{def: } 0, 1, \beta_3, \alpha_2, \beta_2, \beta_1, \alpha_1. \tag{7.30}$$

Let

(7.31) $$\mathscr{P}'(\zeta) = \text{def} = \mathscr{P}(\zeta^{\smile_1} \times \zeta^{\frown}, \zeta^{\smile_1} \times \zeta^{\frown}).$$

By (7.2), (7.30), (7.31) we have

$$\mathscr{P}'(0) = \mathscr{P}(\alpha_1 \times 1, \alpha_2 \times 1) = \alpha_1$$
$$\mathscr{P}'(\beta_2) = \mathscr{P}(1 \times \alpha_2, \beta_1 \times \alpha_2) = 0 \quad \text{(by (7.28))}$$
$$\mathscr{P}'(\beta_1) = \mathscr{P}(\beta_2 \times \alpha_1, 1 \times \alpha_1) = 0 \quad \text{(by (7.29))}$$
$$\mathscr{P}'(\zeta) = 0 \ (\zeta \neq 0, \neq \beta_2, \neq \beta_1), \quad \text{(for simple reasons).}$$

Hence by (7.30),

(7.31.1) $$(\mathscr{P}'(\zeta))^{\frown} = \Delta(\zeta)$$

and case $3a_4$ is proved (theorem 3).

Case $3a_5$, $r = 2$, $k > 3$.

We have

$$B = \{0, 1, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_k\}.$$

Arrange the notation so that

(7.32) $$\beta_{k-2}\alpha_2 = 0.$$

We further observe that

(7.33) $$\beta_i \alpha_1 = 0 \qquad (i = 1, 2, \ldots, k)$$

since from (7.2) we have

$$\beta_i \alpha_1 = \beta_i \mathscr{P}(\alpha_1, \alpha_2) = 0 \quad \text{(prpr of length 3).}$$

Define $\zeta^{\frown}$ by the ordering

(7.34) $$\zeta^{\frown} = \text{def: } 0, 1, \beta_k, \alpha_2, \beta_{k-1}, \beta_{k-2}, \ldots, \beta_2, \beta_1, \alpha_1.$$

Let

(7.35) $$\mathscr{P}'(\zeta) = \text{def} = \mathscr{P}(\zeta^{\smile_1} \times \zeta^{\frown}, \zeta^{\smile_{k+1}} \times \zeta^{\frown}).$$

Then by (7.2) and (7.34),

(7.36)
$$\begin{cases} \mathscr{P}'(0) = \mathscr{P}(\alpha_1 \times 1, \ \alpha_2 \times 1) = \alpha_1 \\ \mathscr{P}'(\beta_k) = \mathscr{P}(1 \times \alpha_2, \ \beta_{k-2} \times \alpha_2) = 0 \quad \text{(by (7.32))} \\ \mathscr{P}'(\beta_1) = \mathscr{P}(\beta_2 \times \alpha_1, \ 1 \times \alpha_1) = 0 \quad \text{(by (7.33))} \\ \mathscr{P}'(\zeta) = 0 \ (\zeta \neq 0, \neq \beta_k, \neq \beta_1), \quad \text{(for simple reasons).} \end{cases}$$

Therefore by (7.34), (7.35)

(7.37) $$(\mathscr{P}'(\zeta))^{\frown} = \Delta(\zeta)$$

and case $3a_5$ is proved (theorem 3).

Case 3b: there exists no $\beta\ (= \beta_1, \beta_2, \ldots, \beta_k)$ and $\alpha\ (= \alpha_2, \alpha_3, \ldots, \alpha_r)$ such that $\beta\alpha = 0$.

Case $3b_4$, $r = 2$.

Here

$$B = \{0, 1, \alpha_1, \alpha_2, \beta_1, \ldots, \beta_k\}.$$

We first prove the

LEMMA b_1. $\beta_i \beta_j = 0 \qquad (i \neq j;\ i, j, = 1, 2, \ldots, k)$.

Proof. Under the prevailing hypotheses,

$$\beta_i \alpha_2 = 0 \quad \text{or} \quad \beta_i \quad \text{or} \quad \alpha_2.$$

This follows since the rank is 2 and we are dealing with type *ii*. We observe however that $\beta_i \alpha_2 \neq 0$, by the hypothesis of case 3 b; also $\beta_i \alpha_2 \neq \alpha_2$, for if $\beta_i \alpha_2 = \alpha_2$ then, from (7.2), we should have

$$\mathscr{P}(\alpha_1, \beta_i \alpha_2) = \alpha_1,$$

a contradiction since the left $= 0$ (prpr of length 3). There remains only the alternative $\beta_i \alpha_2 = \beta_i$. From this we have

$$(\beta_i \alpha_2)\, \beta_j = \beta_i \beta_j. \tag{7.38}$$

For $i \neq j$ the left of (7.38) is 0 (prpr of length 3), whence lemma b_1 follows.

Returning now to the proof of case 3 b_1, let $\zeta^\frown$ be defined by the ordering

$$\zeta^\frown = \text{def: } 0, 1, \alpha_2, \beta_k, \beta_{k-1}, \ldots, \beta_1, \alpha_1. \tag{7.39}$$

Let

$$\mathscr{P}'(\zeta) = \text{def} = \mathscr{P}(\zeta^{\smile_1}, \zeta^{\smile_{k+1}} \times \zeta^\frown). \tag{7.40}$$

By (7.2), (7.39), (7.40) we have

$$\mathscr{P}'(0) = \mathscr{P}(\alpha_1, \alpha_2 \times 1) = \alpha_1$$

$$\mathscr{P}'(\alpha_2) = \mathscr{P}(1, \beta_{k-1} \times \beta_k) = 0 \quad \text{(by lemma } b_1\text{)}$$

$$\mathscr{P}'(\zeta) = 0 \ (\zeta \neq 0, \neq \alpha_2), \quad \text{(for simple reasons)}.$$

Hence

$$\mathscr{P}'(\zeta) = \begin{cases} \alpha_1 & \text{if } \zeta = 0 \\ 0 & \text{if } \zeta \neq 0. \end{cases}$$

Therefore, by (7.39),

$$(\mathscr{P}'(\zeta))^\frown = \Delta(\zeta) \tag{7.41}$$

and case 3 b_1 is proved (theorem 3).

Case 3 b_2, $r > 2$.

We first prove the

LEMMA b_2. *Let $(B, \times)$ be a finite binary algebra satisfying all hypotheses prevailing in case 3 b_2. Let Q be a prpr of length r which (1°) contains at least one β_i $(i = 1, 2, \ldots, k)$, (2°) contains at least one α_j $(j = 1, 2, \ldots, r)$ and (3°) omits at least one α_t $(t = 2, 3, \ldots, r)$. Then $Q = 0$.*

Proof. Suppose Q satisfies (1°) to (3°). Then, since $(B, \times)$ is of type *ii*,

$$Q = 0 \quad \text{or} \quad \alpha \quad \text{or} \quad \beta$$

where α is some α_i appearing in Q, and β is some β_j appearing in Q. However $Q(\ldots, \alpha, \ldots) \neq \alpha$ since from $Q = \alpha$ together with (7.2) we get

$$\mathscr{P}(\alpha_1, \ldots, Q(\ldots, \alpha, \ldots), \ldots, \alpha_r) = \alpha_1,$$

a contradiction since the left side $= 0$ (prpr of length $> r$). But $Q = \beta$ is also impossible, for from $Q = \beta$, on multiplying both sides by an α_t ($t = 2, 3, \ldots, r$) not occurring in Q (see (3°) above) we get

$$Q \times \alpha_t = \beta \times \alpha_t.$$

This is a contradiction since the left side $= 0$ (prpr of length $r+1$) while the right side $\neq 0$ (part of the hypothesis of case 3 b). There remains only the alternative $Q = 0$, which proves lemma b_2.

Returning now to the proof of case 3 b_2, define $\zeta^\frown$ by the ordering

$$\zeta^\frown = \text{def: } 0, 1, \alpha_r, \alpha_{r-1}, \ldots, \alpha_3, \alpha_2, \beta_k, \beta_{k-1}, \ldots, \beta_2, \beta_1, \alpha_1. \tag{7.42}$$

Let

$$\mathscr{P}'(\zeta) = \text{def} = \mathscr{P}(\zeta^\smile, \zeta^{\smile_{s+k}}, \zeta^{\smile_{s+k}}, \ldots, \zeta^{\smile_{r+k}}) \times \zeta^\frown. \tag{7.43}$$

From (7.2), (7.42), (7.43) we have

$$\mathscr{P}'(0) = \mathscr{P}(\alpha_1, \alpha_2, \ldots, \alpha_r) \times 1 = \alpha_1$$

$$\mathscr{P}'(\alpha_r) = \mathscr{P}(1, \beta_{k-1}, \beta_k, \alpha_2, \ldots) \times \alpha_{r-1} = 0 \quad \text{(by lemma } b_2\text{)}$$

$$\mathscr{P}'(\zeta) = 0 (\zeta \neq 0, \neq \alpha_r), \quad \text{(for simple reasons).}$$

Hence

$$\mathscr{P}'(\zeta) = \begin{cases} \alpha_1 & \text{for } \zeta = 0 \\ 0 & \text{for } \zeta \neq 0. \end{cases} \tag{7.44}$$

From (7.42), (7.44) we have

$$(\mathscr{P}'(\zeta))^\frown = \Delta(\zeta) \tag{7.45}$$

and case 3 b_2 is proved (theorem 3). The several cases considered in this section are readily seen to be exhaustive and thus to complete the proof of lemma 7.

8. Rank 1; completion of theorem 2. In the foregoing sections theorem 2 has been proved for all finite binary algebras (1°) of order $n < 4$, and also for all finite binary algebras (2°) of order $n \geqq 4$ and rank $r \geqq 2$. There remains then only the case (3°), order $n \geqq 4$ and rank $r = 1$. Let

$$B = \{0, 1, \alpha_1, \alpha_2, \ldots, \alpha_{n-2}\}$$

be the elements of a finite binary algebra of order $n \geqq 4$ and rank $r = 1$. We remark that a necessary (but not a sufficient) condition for a binary algebra to be of rank 1 is that $\alpha_i \alpha_j = 0$ ($i \neq j$; $i, j = 1, 2, \ldots, n-2$). Hence

(I) The class of finite binary algebras for which all products $\alpha_i \alpha_j = 0$ ($i \neq j$) includes all finite binary algebras of rank 1.

LEMMA 8.1. *Let $(B, \times)$ be a finite binary algebra of order $n>4$, in which the product of any two distinct proper elements $=0$. Then the conclusion of theorem 2 holds.*

Proof. Let $\zeta^\frown$ be defined by an arbitrary but fixed ordering of B, e.g.,

$$\zeta^\frown = \text{def}: \quad 0, 1, \alpha_1, \alpha_2, \ldots, \alpha_{n-2}. \tag{8.1}$$

It is then readily verified that

$$\{(\zeta\zeta^\frown)^\smile(\zeta\zeta^\frown)^{\frown n-3}\}^\frown = \Delta_1(\zeta). \tag{8.2}$$

Lemma 8.1 is thus proved, by theorem 3.

In view of the above observation (I) we immediately have the corollary

LEMMA 8.2. *Let $(B, \times)$ be a finite binary algebra of order $n>4$ and rank $r=1$. Then the conclusion of theorem 2 holds.*

Finally there remains only the case of: order $n=4$ and rank $r=1$. Let $(B, \times)$ be of order 4 and rank 1,

$$B = \{0, 1, \alpha_1, \alpha_2\}.$$

In any event we must have

$$\alpha_1\alpha_2 = \alpha_2\alpha_1 = 0$$

otherwise the rank would be 2. What about α_1^2 and α_2^2? We list the possible values of α_1^2, α_2^2 in tabular form below. Some of the corresponding binary algebras are of rank 2, and for such theorem 2 has already been proved; for those of rank 1 an expression for $\Delta_1(\zeta)$ or for $\delta_1(\zeta)$ will be given, from which the conclusion of theorem 2 follows, on use of theorem 3. A single exception is formed by the last case listed, which corresponds to $(H_4, \times)$ of theorem 2; we shall show that this cannot be functionally completed by any choice of $\zeta^\frown$ (see however § 10).

α_1^2	α_2^2	rank	
1	arbitrary	2	$(\alpha_1^2)\,\alpha_2 \neq 0$
arbitrary	1	2	$(\alpha_2^2)\,\alpha_1 \neq 0$
α_1	α_1	2	$(\alpha_1^2)\,(\alpha_2^2) \neq 0$
α_2	α_2	2	$(\alpha_1^2)\,(\alpha_2^2) \neq 0$
α_2	α_1	2	$\alpha_1\,(\alpha_2^2) \neq 0$
0	0	1	define $\zeta^\frown$ by the ordering 0, 1, α_1, α_2; then $\zeta^2 = \delta_1(\zeta)$
0	α_2	1	define $\zeta^\frown$ by the ordering 0, 1, α_2, α_1; then $(\zeta^2 \times \zeta^{\frown 3})^\frown = \Delta_1(\zeta)$
α_1	0	1	interchange α_1, α_2 in preceding case
0	α_1	1	define $\zeta^\frown$ by the ordering 0, 1, α_2, α_1; then $(\zeta\zeta^\frown)^{2\frown} = \Delta_1(\zeta)$
α_2	0	1	interchange α_1, α_2 in preceding case
α_1	α_2	1	corresponds to $(H_4, \times)$; see below

Consider finally the last case above,

$$(H_4, \times); \quad \{0, 1, \alpha_1, \alpha_2\}; \quad \alpha_1\alpha_2 = \alpha_2\alpha_1 = 0, \quad \alpha_1^2 = \alpha_1, \ \alpha_2^2 = \alpha_2.$$

The only cyclic permutations of H_4 with $0^\frown = 1$ are of course

$$\zeta^\frown = \text{def}: \ 0, 1, \alpha_1, \alpha_2 \quad \text{or else} \quad \zeta^\frown = \text{def}: \ 0, 1, \alpha_2, \alpha_1.$$

We observe that $\{0, \alpha_1\}, \{1, \alpha_2\}$ form the residue classes (or congruence classes) of a proper ideal of the algebra $(H_4, \times, \frown)$ if $\zeta^\frown$ is defined by the first ordering above, while dually $\{0, \alpha_2\}, \{1, \alpha_1\}$ are the residue classes of an ideal if $\zeta^\frown$ is defined by the second ordering above. We now appeal to the following result taken from [1],

THEOREM 4. *A primal algebra $\mathfrak{P}$ is simple, i.e., possesses no ideal other than the universe ideal ($=\mathfrak{P}$ itself).*

Accordingly, from the remarks above, $(H_4, \times)$ cannot be converted into a primal algebra, and we have proved

LEMMA 8.3. *Let $(B, \times)$ be a binary algebra of order 4 different from the algebra $(H_4, \times)$ of theorem 2. Then the conclusion of theorem 2 holds. The algebra $(H_4, \times)$ cannot be completed to a primal frame by any choice of $\zeta^\frown$.*

The set of all foregoing lemmas exhaust all possible cases, and complete the proof of theorem 2.

9. Special binary algebras. We have shown that any finite binary frame, not isomorphic with $(H_4, \times)$ may be "completed" to a primal frame in at least one — but usually not in an arbitrary — way. Thus, for example, $(R_4, \times)$, the multiplicative residues mod 4, $R = \{0, 1, 2, 3\}$, becomes a primal frame $(R, \times, \frown)$ if $\zeta^\frown$ is defined by the ordering 0, 1, 3, 2, while it is not a primal algebra if $\zeta^\frown$ is defined by the "natural" ordering 0, 1, 2, 3; this last follows by a simple application of theorem 4.

There are however extensive and important classes of binary algebras which, unlike $(R_4, \times)$ above, are "arbitrarily" completable. We have

THEOREM 5. *Let $(B, \times)$ be a finite binary algebra, of order n, and which satisfies at least one of the following:*

(1′) *There exists a product $\mathscr{P}$ of length $n-2$ such that $\mathscr{P} = 1$.*

(1″) *For each $\mu \in B$, with the possible exception of $\mu = 1$, there exists a product $\mathscr{P}$ ($= \mathscr{P}_\mu$) of length $n-2$ such that $\mathscr{P}_\mu = \mu$.*

Then if $0, 1, \beta_1, \beta_2, \ldots, \beta_{n-2}$ is an arbitrary ordering of B and if $\zeta^\frown$ is correspondingly defined (i.e., $0^\frown = 1, 1^\frown = \beta_1, \ldots, (\beta_{n-2})^\frown = 0$), $(B, \times, \frown)$ is a functionally strictly complete algebra (primal frame).

It is noted that the condition (1′) implies (1″). The proof of theorem 5 follows readily from the proofs of lemma 6, case 1 and lemma 7, case 1; we omit the details.

REMARK. The satisfaction of either (1′) or (1″) of theorem 5 implies of course that $(B, \times)$ is of (maximal) rank $n-2$. However it must not be inferred that this implies the absence of zero divisors. Thus, in $(K_4, \times)$ (§ 4), $\alpha\beta = 0$ but $(K, \times)$ is of rank 2, and indeed (1′) is satisfied with $\mathscr{P}(\alpha, \beta) = (\alpha^2)(\beta^2) = 1$.

Of the many special cases of theorem 5 we call attention to but one (see also [1]). Let $(G', \times)$ be a group with its identity written as 1. By $(G, \times)$, called a "group with null", we mean the binary algebra in which $G = G'$ augmented by a null element, 0:

$$\zeta \times \eta \text{ in } G = \zeta \times \eta \text{ of } G', \text{ for } \zeta, \eta \in G'; \quad 0 \times \zeta = \zeta \times 0 = 0 \quad (\zeta \in G).$$

THEOREM 6. *Let $(G, \times)$ be a finite group with null and let $\zeta^\frown$ be defined corresponding to an arbitrary but fixed ordering of its elements. Then $(G, \times, \frown)$ is a primal frame (="group-frame").*

Proof. Let $G = \{0, 1, \omega_1, \omega_2, \ldots, \omega_{n-1}\}$, and take

$$\mathscr{P}(\omega_1, \omega_2, \ldots, \omega_{A-1}) = \omega_1^n \omega_2^n \ldots \omega_{n-1}^n.$$

Since $\mathscr{P} = 1$ the proof follows from theorem 5. It is noted that the finite Galois fields $(F_{p^k}, \times, \frown)$ form a special case of theorem 6.

10. Remark on $(H_4, \times)$. In theorem 2 the exceptional role of the algebra $(H_4, \times)$ disappears if we abandon our normalization condition $0^\frown = 1$ for the cyclic permutation $\zeta^\frown$. In particular, if $\zeta^\frown$ is defined in H_4 by

$$\zeta^\frown = \text{def} = 0, \alpha, 1, \beta$$

the algebra $(H_4, \times, \frown)$ is functionally strictly complete. Indeed it is readily verified that the characteristic function $\delta_\beta(\zeta)$ is here given by the expression

$$\delta_\beta(\zeta) = \varphi(\zeta) \times \psi(\zeta)$$

where

$$\varphi(\zeta) = \{(\zeta\zeta^\frown\zeta^{\frown 2}\zeta^{\frown 3})^\frown(\zeta\zeta^\smile)^\frown\}^{\frown 2}$$

$$\psi(\zeta) = \{(\zeta\zeta^\frown\zeta^{\frown 2}\zeta^{\frown 3})^\frown \times \varphi(\zeta)\}^\frown.$$

Bibliography

[1] FOSTER, A. L.: Generalized "Boolean" theory of universal algebras. Part I: Subdirect sums and normal representation theorem. Math. Z. **58**, 306–336 (1953). — [2] FOSTER, A. L.: Generalized "Boolean" theory etc. Part II: Identities and subdirect sums of functionally complete algebras. Math. Z. **59**, 191–199 (1953). — [3] FOSTER, A. L.: The identities of–and unique subdirect factorization within–classes of universal algebras. Math. Z. **62**, 171–188 (1955). — [4] FOSTER, A. L.: On n-altiy theories in rings and their logical algebras etc. Amer. J. Math. **72**, 101–123 (1950). — [5] MCCOY, N. H., and DEANE MONTGOMERY: A representation of generalized Boolean rings. Duke math. J. **3**, 455–459 (1937). — [6] ROSENBLOOM, P.: The elements of math. logic. New York: Dover 1950. — [7] WADE, L. I.: Post algebras and rings. Duke math. J **12**, 389–395 (1945).

Dept. of Math., University of California, Berkeley 4, Cal. (U.S.A.)

(Eingegangen am 11. August 1958)

GAIER, D.
Math. Zeitschr. 71, 83–88 (1959)

Über ein Extremalproblem der konformen Abbildung

Von
DIETER GAIER

In den letzten zehn Jahren hat der von AHLFORS und BEURLING in verschiedenen Arbeiten (vgl. [1], [2], [3], [5]) eingeführte Begriff der extremalen Länge $\lambda(\Gamma)$ einer Kurvenfamilie Γ immer mehr an Bedeutung für die Theorie der konformen Abbildung und Anwendungen in der Funktionentheorie gewonnen. Eine zusammenfassende Darstellung findet man bei JENKINS [6].

Wir betrachten zum Beispiel als zugrunde gelegtes Gebiet G den Kreisring $1<|z|<M$, und als Γ die Klasse aller rektifizierbaren Kurven $\gamma<G$, die $|z|=1$ umschließen. Ferner sei $\varrho(z)\geqq 0$ in G so erklärt, daß

$$(1) \qquad A_\varrho(G)=\iint_G \varrho^2\,dx\,dy \quad \text{und} \quad L_\varrho(\Gamma)=\inf_{\gamma\in\Gamma}\int_\gamma \varrho\,|dz|$$

im Lebesgueschen Sinne existieren und nicht gleichzeitig 0 oder ∞ sind. Dann nennt man

$$(2) \qquad \lambda(\Gamma)=\sup_\varrho \frac{[L_\varrho(\Gamma)]^2}{A_\varrho(G)}$$

die extremale Länge der Kurvenfamilie Γ. In diesem Falle des Kreisrings ergibt sie sich zu $\lambda(\Gamma)=2\pi/\log M$, und die extremale Metrik ist $\varrho_0(z)|dz|=\frac{|dz|}{|z|}$, so daß diese also hier nicht konform äquivalent zur euklidischen Metrik ist.

Obwohl für die Anwendungen des Begriffs der extremalen Länge gerade die Allgemeinheit der zugelassenen Funktionen $\varrho(z)$ wichtig ist, kann man sich daher die Frage vorlegen, wie groß $\lambda(\Gamma)$ wird, wenn zur Konkurrenz in (2) nur Funktionen $\varrho(z)$ von der Form $\varrho(z)=|f'(z)|$ zugelassen werden, wobei $f(z)$ in G eindeutig-regulär und schlicht ist; auch dann würde $\lambda(\Gamma)$ eine konforme Invariante für G. Anders ausgedrückt: Es bezeichne $\mathfrak{R}_M$ die Klasse aller endlichen zweifach zusammenhängenden Gebiete G vom Modul M (die sich also konform auf $1<|z|<M$ abbilden lassen); für jedes $G\in\mathfrak{R}_M$ sei $A(G)$ die (innere) Fläche von G, und $L(G)$ bezeichne die untere Grenze der Längen aller rektifizierbaren Kurven in G, welche die innere Randkomponente von G umschließen. Gesucht ist $\sup_{G\in\mathfrak{R}_M}\frac{[L(G)]^2}{A(G)}$.

Der Wert dieser Größe ist uns nicht bekannt, dagegen werden wir in dieser Note folgendes damit eng zusammenhängende, einfachere Problem lösen. Wieder bezeichne $\mathfrak{R}_M$ die Klasse aller endlichen zweifach zusammenhängenden Gebiete G der w-Ebene ($w=u+iv$) vom Modul M; für jedes $G\in\mathfrak{R}_M$ sei $A(G)$ jetzt die (innere) Fläche innerhalb der äußeren Randkomponente, und $D(G)$ bezeichne die untere Grenze der Durchmesser aller Kurven in G, welche die

innere Randkomponente von G umschließen. $D(G)$ ist also der Durchmesser der inneren Berandung von G. Gesucht wird eine Abschätzung für $\frac{[D(G)]^2}{A(G)}$. Eine solche Abschätzung ist auch bei der Untersuchung der Konvergenzgeschwindigkeit von Iterationsverfahren der konformen Abbildung nützlich, worauf wir demnächst zurückkommen werden.

Eine Antwort auf das gestellte Problem gibt folgender

SATZ. *Mit den oben angegebenen Bezeichnungen ist*

$$\sup_{G\in\mathfrak{R}_M}\frac{[D(G)]^2}{A(G)}=\frac{16}{\pi}\cdot\sum_{\substack{\nu=1\\ \text{ungerade }\nu}}^{\infty}\frac{M^{2\nu}}{\nu(M^{4\nu}-1)}. \tag{3}$$

Bildet $w=f_0(z)$ den gespiegelten Kreisring $1<|z|<M^2$ auf einen zur u-Achse symmetrischen und längs $-1\leqq u\leqq+1$ geschlitzten Parallelstreifen konform ab, so ist ein Extremalgebiet G_0 unseres Problems von C_1: $-1\leqq u\leqq+1$ sowie von $f_0(|z|=M)$ berandet. Jedes weitere Extremalgebiet ergibt sich aus G_0 durch eine ganze lineare Abbildung.

Für die rechte Seite von (3) und für die Extremalabbildung $f_0(z)$ werden wir unten einen expliziten Ausdruck angeben.

Unser Ergebnis kann auch als *isoperimetrische Ungleichung für Kondensatoren* ausgedrückt werden; vgl. hierzu vor allem [9]. Gegeben sei ein zylindrischer Kondensator, dessen Querschnitt zwei Kurven C_1 und C_2 bilden, deren innere einen Durchmesser D habe, während die äußere eine Fläche A umschließe. *Ist dann c die Kapazität des Kondensators pro Höheneinheit, so gilt*

$$\frac{D^2}{A}\leqq\frac{16}{\pi}\cdot\sum_{\substack{\nu=1\\ \text{ungerade }\nu}}^{\infty}\frac{e^{\nu/c}}{\nu(e^{2\nu/c}-1)}.$$

Das Gleichheitszeichen wird für den über G_0 errichteten Kondensator angenommen, wobei $M=e^{1/2c}$ zu wählen ist. Dies ergibt sich aus (3), wenn man die eben angegebene Relation zwischen der Kapazität des Zylinders pro Höheneinheit und dem Modul des Querschnitts verwendet.

Schließlich wollen wir noch folgendes bemerken. Ist G das Innere der längs $-1\leqq u\leqq+1$ geschlitzten Ellipse mit Brennpunkten in ±1, so daß G den Modul M hat, dann ergibt sich für dieses Gebiet

$$\frac{[D(G)]^2}{A(G)}=\frac{16}{\pi}\cdot\frac{M^2}{M^4-1},$$

also der erste Term der rechten Seite in (3). Unser Gebiet G kann daher in gewissem Sinne als erste Annäherung für G_0 betrachtet werden.

Der Beweis unseres Satzes stützt sich auf die Methode der zirkularen Symmetrisierung und auf Ergebnisse von GARABEDIAN und SCHIFFER [4].

a. Neben der Klasse $\mathfrak{R}_M$ betrachten wir diejenige Unterklasse $\mathfrak{L}_M$ von $\mathfrak{R}_M$, die alle $G\in\mathfrak{R}_M$ enthält, bei denen die innere Berandung C_1 aus der Strecke

$-1 \leqq u \leqq +1$ besteht, während die äußere Berandung diese Strecke umschließt und zur u-Achse symmetrisch ist. Wir setzen

$$\alpha = \sup_{G \in \mathfrak{R}_M} \frac{[D(G)]^2}{A(G)} \quad \text{und} \quad \beta = \sup_{G \in \mathfrak{L}_M} \frac{[D(G)]^2}{A(G)}.$$

Dann ist sicher $\alpha \geqq \beta$, und wir zeigen im ersten Beweisschritt jetzt $\beta \geqq \alpha$. Dazu konstruieren wir zu jedem $G \in \mathfrak{R}_M$, welches sich durch keine ganze lineare Abbildung in ein Gebiet $\in \mathfrak{L}_M$ überführen läßt, ein $G^* \in \mathfrak{L}_M$ mit

$$\frac{[D(G)]^2}{A(G)} < \frac{[D(G^*)]^2}{A(G^*)} \tag{4}$$

wie folgt. Wir können annehmen, daß der Durchmesser der Innenkurve von G in ± 1 angenommen wird; eine ganze lineare Abbildung läßt ja D^2/A invariant. In einer ersten Operation werde G bezüglich $w = -1$ und $u \geqq -1$ zirkular symmetrisiert; vgl. dazu [9], S. 193ff. und [6], S. 130ff. Dabei geht G in G' über, welches einen Modul $M' \geqq M$ besitzt, während

$$D(G') = D(G) \quad \text{und} \quad A(G') = A(G)$$

gelten.

Für den Fall $M' > M$, der genau dann eintritt, wenn G' nicht durch Drehung aus G hervorgeht (vgl. [6], S. 136), stellen wir zweitens aus G' ein G'' vom Modul M her. Bezeichnet G'_t das längs $1 \leqq u' \leqq 1+t$ $(t \geqq 0)$ geschlitzte Gebiet G', und ist M'_t sein mit t stetig variierender Modul, so ist $M'_0 = M' > M$, während $M'_t \to 1 < M$ strebt, falls $1+t$ gegen einen Punkt des Außenrandes von G' strebt. Also gibt es ein t_0 mit $M'_{t_0} = M$, und wir setzen $G'_{t_0} = G''$. Damit haben wir ein zirkularsymmetrisches G'' mit Modul $M'' = M$, $D(G'') > D(G)$, $A(G'') = A(G)$ erhalten. Im Falle $M' = M$ unterbleibt diese zweite Operation, d. h. wir setzen dann $G'' = G'$.

Ist nun G'' von C_1'' und C_2'' berandet, und ist C_1'' von der Strecke $-1 \leqq u'' \leqq 1+t_0$ verschieden, so stellen wir drittens aus G'' ein G''' dadurch her, daß wir C_1'' durch jene Strecke ersetzen, während der Außenrand C_2'' von G'' erhalten bleibt. Da G''' das Gebiet G'' echt enthält, ergibt diese Operation

$$D(G''') = D(G''), \quad A(G''') = A(G''), \quad M''' > M'' = M,$$

und durch Einführen eines Schlitzes in G''' wie beim zweiten Schritt stellen wir ein $G^{(4)}$ her mit

$$D(G^{(4)}) > D(G'''), \quad A(G^{(4)}) = A(G'''), \quad M^{(4)} = M.$$

In dem Fall, daß jedoch C_1'' ein geradliniger Schlitz ist, unterbleiben diese beiden Operationen, d. h. wir setzen dann $G^{(4)} = G''$.

Eine Streckung von $w^{(4)} = -1$ aus ergibt schließlich in jedem Falle ein G^* mit

$$D(G^*) = D(G), \quad A(G^*) \leqq A(G), \quad M^* = M,$$

das überdies zur u-Achse symmetrisch ist und als Innenberandung die Strecke $-1 \leqq u^* \leqq +1$ hat. Unser konstruiertes Gebiet G^* ist somit $\in \mathfrak{L}_M$ und genügt

$$\frac{[D(G^*)]^2}{A(G^*)} \geqq \frac{[D(G)]^2}{A(G)} \tag{5}$$

Steht hierin das Gleichheitszeichen, so unterbleiben notwendig die vierte, dritte und zweite Operation, d.h. $G' = G^*$ und $M' = M$ muß gelten, also ist schon $G \in \mathfrak{L}_M$. Das beliebig vorgegebene Ausgangsgebiet läßt sich somit in diesem Falle durch eine ganze lineare Abbildung in ein $G^* \in \mathfrak{L}_M$ überführen.

Damit ist unsere Aussage (4) bewiesen. Es genügt nunmehr, unser Extremalproblem in der Unterklasse $\mathfrak{L}_M$ zu betrachten; außerdem ist durch (4) gezeigt, daß etwaige Extremalgebiete aus $\mathfrak{R}_M$ notwendig schon in $\mathfrak{L}_M$ liegen oder wenigstens durch eine ganze lineare Abbildung aus solchen hervorgehen.

b. Nun ist $D(G) = 2$ $(G \in \mathfrak{L}_M)$; also ist $A(G)$ nach unten abzuschätzen, was im zweiten Beweisschritt geschehen soll. Hierzu sei $w = f(z)$ die konforme Abbildung von $1 < |z| < M$ auf G, wobei $|z| = 1$ in $-1 \leqq u \leqq +1$ übergehe und $f(1) = 1$, $f(-1) = -1$ normiert sei. Nach dem Spiegelungsprinzip ist $f(z)$ sogar in $M^{-1} < |z| < M$ regulär, und das Bild dieses Kreisrings unter der Abbildung $w = f(z)$ hat die Fläche $2A(G)$.

Nun bezeichne $\omega = P(z)$ diejenige konforme Abbildung von $M^{-1} < |z| < M$ auf ein zu $\omega = 0$ konzentrisches Kreisschlitzgebiet, für die $P(-1) = 0$ und $P(+1) = \infty$ ist, und $\omega = Q(z)$ sei die analoge Radialschlitzabbildung. Normieren wir $P(z)$ und $Q(z)$ durch die Vorschrift, daß das Residuum in $z = +1$ jeweils gleich 1 sein soll, so liegen diese Funktionen eindeutig fest. Unter Verwendung der Greenschen Formel, des Residuensatzes und der Schwarzschen Ungleichung findet man dann unschwer für jede in $M^{-1} < |z| < M$ reguläre Funktion $F(z)$ (vgl. [4], S. 190ff. und [3], S. 113–114)

$$|F(-1) - F(1)|^2 \leqq \frac{1}{2\pi} \log \frac{P'(-1)}{Q'(-1)} \cdot \iint\limits_{M^{-1} < |z| < M} |F'(z)|^2 \, dx \, dy, \tag{6}$$

und Gleichheit besteht genau für Funktionen

$$F_0(z) = L_1 \cdot \log \frac{P(z)}{Q(z)} + L_2 \qquad (L_1, L_2 \text{ konstant}) \tag{7}$$

Setzt man insbesondere $F(z) = f(z)$ in (6) ein, so ergibt sich

$$2A(G) \geqq 8\pi / \log \frac{P'(-1)}{Q'(-1)}, \tag{8}$$

insgesamt also

$$\frac{[D(G)]^2}{A(G)} \leqq \frac{1}{\pi} \cdot \log \frac{P'(-1)}{Q'(-1)} \qquad \text{für jedes } G \in \mathfrak{L}_M \tag{9}$$

Wir zeigen nun in **c**, daß bei geeigneter Wahl von L_1 und L_2 in (7) $F_0(z)$ als Bild von $1 < |z| < M$ ein Gebiet $G_0 \in \mathfrak{L}_M$ liefert, daß also $F_0(z)$ ein $f(z)$ wird. Dann erscheint für G_0 in (9) das Gleichheitszeichen, und man ersieht aus obigem wegen der Normierung von C_1, daß es auch nur in diesem Fall stehen kann.

c. Wir bestimmen zunächst die Schlitzabbildungen $P(z)$ und $Q(z)$ von $M^{-1} < |z| < M$ und können uns dabei auf Darstellungen dieser Funktionen

durch ein *ON*-System beziehen ([7], S. 169–170 und [8], S. 376–377). Ist $\{u_\nu(z)\}$ irgendein vollständiges und durch

$$\frac{1}{2i}\int_{|z|=M,\ |z|=M^{-1}} u_\nu'(z)\,\overline{u_\mu(z)}\,dz = \delta_{\nu\mu}$$

orthonormiertes Funktionensystem für $M^{-1}<|z|<M$, so ergibt eine leichte Rechnung in unserem Falle ([8], S. 377)

$$\log\frac{P(z)}{Q(z)} = 2\pi\cdot\sum_\nu\{[\overline{u_\nu(-1)} - \overline{u_\nu(1)}]\cdot u_\nu(z)\} + L_3,$$

wobei die Konstante L_3 durch

$$\log\frac{P(z)}{Q(z)}\Big|_{(z=1)} = 0 \tag{10}$$

festgelegt ist. Wählt man als Funktionensystem

$$u_\nu(z) = \frac{z^\nu}{\sqrt{\pi\nu(M^{2\nu}-M^{-2\nu})}} \qquad (\nu = \pm 1, \pm 2, \ldots),$$

so wird

$$\log\frac{P(z)}{Q(z)} = -4\cdot\sum_{\substack{\nu=-\infty\\ \text{ungerade }\nu}}^{+\infty}\frac{z^\nu}{\nu(M^{2\nu}-M^{-2\nu})} + L_3,$$

und unter Berücksichtigung von (10) also schließlich

$$\log\frac{P(z)}{Q(z)} = 4\cdot\sum_{\substack{\nu=-\infty\\ \text{ungerade }\nu}}^{+\infty}\frac{1-z^\nu}{\nu(M^{2\nu}-M^{-2\nu})} = 4\cdot\sum_{\substack{\nu=1\\ \text{ungerade }\nu}}^{\infty}\frac{2-z^\nu-z^{-\nu}}{\nu(M^{2\nu}-M^{-2\nu})}. \tag{11}$$

Setzt man hierin $z=-1$ ein, so erhält man aus (9) die Abschätzung

$$\frac{[D(G)]^2}{A(G)} \leqq \frac{16}{\pi}\cdot\sum_{\substack{\nu=1\\ \text{ungerade }\nu}}^{\infty}\frac{M^{2\nu}}{\nu(M^{4\nu}-1)} \qquad (G\in\mathfrak{L}_M). \tag{12}$$

Um zu einer expliziten Darstellung von $\log\frac{P(z)}{Q(z)}$ durch elliptische Funktionen zu kommen, setzen wir $z=e^{2\pi i v}$, $2-z^\nu-z^{-\nu}=4\sin^2\pi\nu v$, also

$$\log\frac{P(z)}{Q(z)} = g(v) = 16\cdot\sum_{\substack{\nu=1\\ \text{ungerade }\nu}}^{\infty}\frac{M^{-2\nu}}{\nu(1-M^{-4\nu})}\sin^2\pi\nu v. \tag{13}$$

Nun ist in den gebräuchlichen Bezeichnungen (vgl. [11], S. 160 und 172)

$$\log\left[\frac{\vartheta_3(v)\,\vartheta_4(0)}{\vartheta_4(v)\,\vartheta_3(0)}\right] = -8\cdot\sum_{\substack{\nu=1\\ \text{ungerade }\nu}}^{\infty}\frac{q^\nu}{\nu(1-q^{2\nu})}\sin^2\pi\nu v = \log d\,n(2Kv;k),$$

wobei $q=e^{-\pi K'/K}$, $k=k(q)$ bekannte Größen sind. Aus (13) erhalten wir daher

$$\log\frac{P(z)}{Q(z)} = -2\log d\,n(2Kv;k), \tag{14}$$

und dabei ist der Jacobische Parameter $q=M^{-2}$ zu wählen. Diese Funktion bildet $1<|z|<M^2$ auf $|\Im(w)|<\pi$ ab, versehen mit einem Schlitz von $w=0$

bis $w = -2\log k' > 0$, und $|z| = 1$ wird in den Schlitz abgebildet. Also kann man L_1, L_2 in (7) so wählen, daß das Bild von $1 < |z| < M$ unter $w = F_0(z)$ ein $G_0 \in \mathfrak{L}_M$ wird.

Aus (14) gewinnen wir auch eine explizite Darstellung der rechten Seiten von (9) oder (12). Diese sind

$$\frac{1}{\pi} \cdot \log \frac{P'(-1)}{Q'(-1)} = \frac{1}{\pi} \cdot \log \frac{P(z)}{Q(z)}\Big|_{(z=-1)} = -\frac{2}{\pi} \log d\,n\,(K; k) = -\frac{2}{\pi} \log k',$$

wobei k' der zu $k = k(M^{-2})$ komplementäre Modul ist. Daraus entnehmen wir außerdem, daß sich die rechte Seite von (3) für $M \to 1$ asymptotisch wie $\frac{\pi}{2} \cdot \frac{1}{M-1}$ verhält, und eine einfache Überlegung zeigt, daß für alle $M > 1$ stets

$$\frac{[D(G)]^2}{A(G)} < \frac{\pi}{2\log M} \qquad (G \in \mathfrak{R}_M)$$

gilt.

Wir beschließen die Arbeit mit einer kleinen Tabelle für die rechte Seite von (3); dabei wurden für $M^{-2} \leqq 0{,}55$ die Tafeln [10] verwendet, während die anderen Werte von der hiesigen Z 22 berechnet wurden.

$M^{-2} =$	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
$-\frac{2}{\pi}\log k' =$	0,2555	0,5161	0,7873	1,075	1,386	1,727	2,110	2,546	3,052	3,650

$M^{-2} =$	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	1
$-\frac{2}{\pi}\log k' =$	4,372	5,267	6,410	7,925	10,038	13,196	18,448	28,935	60,365	∞

Literatur

[1] AHLFORS, L. V.: Conformal Mapping. Vorlesungsmanuskript vom Oklahoma A. und M. College 1951. — [2] AHLFORS, L. V., et A. BEURLING: Invariants conformes et problèmes extrémaux. C. R. Dixième Congrès Math. Scand. 1946. Kopenhagen 1947. — [3] AHLFORS, L. V., and A. BEURLING: Conformal invariants and function-theoretic null-sets. Acta Math. **83**, 101–129 (1950). — [4] GARABEDIAN, P. R., and M. SCHIFFER: Identities in the theory of conformal mapping. Trans. Amer. Math. Soc. **65**, 187–238 (1949). — [5] HERSCH, J.: Longueurs extrémales et théorie des fonctions. Comm. Math. Helv. **29**, 301–337 (1955). — [6] JENKINS, J. A.: Univalent Functions and Conformal Mapping. Berlin-Göttingen-Heidelberg 1958. — [7] NEHARI, Z.: The kernel function and canonical conformal maps. Duke Math. J. **16**, 165–178 (1949). — [8] NEHARI, Z.: Conformal Mapping. New York-Toronto-London 1952. — [9] PÓLYA, G., and G. SZEGÖ: Isoperimetric Inequalities in Mathematical Physics. Princeton 1951. — [10] SCHULER, M., u. H. GEBELEIN: Acht- und neunstellige Tabellen zu den elliptischen Funktionen. Berlin-Göttingen-Heidelberg 1955. — [11] TRICOMI, F., u. M. KRAFFT: Elliptische Funktionen. Leipzig 1948.

Math. Inst. A der Technischen Hochschule, Stuttgart, Keplerstraße 10

(Eingegangen am 2. September 1958)

REMBS, E.
Math. Zeitschr. 71, 89–93 (1959)

Über die Verbiegbarkeit der Rinnen

ENRICO BOMPIANI zum 70. Geburtstag gewidmet

Von

EDUARD REMBS

1. Als Rinne R wird eine mindestens dreimal differenzierbare Fläche mit folgenden Eigenschaften bezeichnet: Sie enthält eine ebene Kurve K in einer Ebene E, liegt ganz auf der einen Seite der Ebene E und wird von E längs K berührt. Die Berührung soll von erster Ordnung sein. K trenne auf R ein Gebiet positiver Gaußscher Krümmung von einem solchen mit negativer Krümmung. K ist also parabolische Kurve auf R. R kann beliebig schmal sein und die Fläche des sphärischen Bildes darf beliebig klein $\neq 0$ sein. Wir nennen die Rinne konvex, wenn K konvex, und geschlossen, wenn K geschlossen ist. Die „geschlossene Rinne" ist also keine geschlossene Fläche. Die beiden Teile, der Teil positiver Krümmung und der negativer Krümmung mit Einschluß beide Male von K, werden Halbrinnen genannt.

Das einfachste Beispiel einer geschlossenen konvexen Rinne erhält man, wenn man einen Torus durch eine Ebene E_1 schneidet, die parallel zu den Ebenen seiner parabolischen Kurven ist und einen beliebig kleinen Abstand von einer dieser Ebenen E_2 hat. Das abgeschnittene Torusstück zwischen E_1 und E_2 ist eine Rotationsrinne.

Die geschlossenen konvexen Rinnen und Halbrinnen sind „starr von zweiter Stufe" [1]. Das bedeutet: Wenn man die Forderung stellt, die Rinne oder Halbrinne $\mathfrak{x}(u, v)$ solle durch analytische Verbiegung in eine auch dreimal differenzierbare Fläche

$$\bar{\mathfrak{x}} = \mathfrak{x} + t\mathfrak{z} + t^2\mathfrak{z}^{(2)} + \cdots$$

übergehen, so daß also $d\bar{\mathfrak{x}}^2 = d\mathfrak{x}^2$, oder $d\mathfrak{x}\, d\mathfrak{z} = 0$ und $2d\mathfrak{x}\, d\mathfrak{z}^{(2)} + d\mathfrak{z}^2 = 0$ gilt, so kann man folgern, daß $\mathfrak{z}$ eine triviale Verbiegung ist, d.h. die Form $\mathfrak{z} = (\mathfrak{a} \times \mathfrak{x}) + \mathfrak{b}$ einer infinitesimalen Bewegung haben muß, während das aus $d\mathfrak{x}\, d\mathfrak{z} = 0$ allein noch nicht geschlossen werden kann.

Die naheliegenden Fragen der endlichen Verbiegbarkeit der Rinnen und der eindeutigen Bestimmtheit durch das Linienelement sind bisher von keiner Seite gestellt worden. Es zeigt sich aber, daß man sie großenteils mit einfachsten Mitteln beantworten kann, vielleicht mit etwas anderem Ergebnis, als man vermuten konnte. Indem wir zu Rinnen isometrische Flächen untersuchen, soll gefordert werden, daß auch sie mindestens dreimal differenzierbar sind.

2. SATZ 1. *Zwei analytische isometrische Rinnen sind kongruent (oder spiegelbildlich). Dasselbe gilt für Halbrinnen.*

Die beiden parabolischen Kurven der Rinnen haben nämlich wegen der Isometrie in entsprechenden Punkten gleiche geodätische Krümmungen. Und da die Flächenstreifen längs der Kurven eben sind, stimmen die geodätischen Krümmungen mit den gewöhnlichen Kurvenkrümmungen überein. Also sind die Kurven kongruent. Wegen der Gleichheit der Gaußschen Krümmungen der beiden Flächen $L_1N_1-M_1^2=L_2N_2-M_2^2$ gilt nun für die Differenzen $l=L_1-L_2$, $m=M_1-M_2$, $n=N_1-N_2$ die Gleichung

$$(N_1+N_2)\,l-2(M_1+M_2)\,m+(L_1+L_2)\,n=0,$$

wie man durch Ausrechnen bestätigt. Infolge der Linearität der Codazzischen Gleichungen gelten diese auch für l, m, n. Indem man beides berücksichtigt, kann man Schritt für Schritt beweisen, daß alle Ableitungen beliebig hoher Ordnung von l, m, n auf der parabolischen Kurve verschwinden. Wegen der genauen Durchführung unter einem etwas anderen Gesichtspunkt vgl. A. D. ALEXANDROW [2]. Die Schlußweise ist übrigens analog zu der in [1] S. 120. Aus Satz 1 folgt noch:

SATZ 2. *Eine analytische Rinne kann nicht in eine andere Rinne verbogen werden.*

3. Wir wollen nun als Biegungsflächen von Rinnen auch Nichtrinnen zulassen, d.h. Flächen, bei denen die Tangentialebenen längs der parabolischen Kurve nicht überall mit ein und derselben Ebene zusammenfallen. Über sie sei vorausgeschickt der

HILFSSATZ. *Wenn eine parabolische Kurve auf einer Fläche ein Gebiet positiver und ein Gebiet negativer Krümmung voneinander trennt und die Tangentialebenen der Fläche längs der Kurve nicht mit ein und derselben Ebene zusammenfallen, durchsetzt die von den Tangentialebenen eingehüllte Fläche die gegebene Fläche, so daß die beiden Teile der Fläche mit verschiedenem Vorzeichen der Gaußschen Krümmung auf verschiedenen Seiten der eingehüllten Fläche liegen.*

BEWEIS. Die parabolische Kurve sei die Parameterkurve $u=0$. Größen auf ihr werden durch angehängte 0 gekennzeichnet. Wir schneiden die Fläche mit den Ebenen durch die Punkte $\mathfrak{x}_0$ der Kurve und den Stellungsvektoren $\xi_v^0/|\xi_v^0|$. Die Schnittkurven wählen wir als Kurven $v=\text{const}$; als Kurven $u=\text{const}$, zu denen also auch die parabolische Kurve gehört, nehmen wir die Kurvenschar, die mit den Kurven $v=\text{const}$ zusammen ein konjugiertes Netz bildet. Nach Voraussetzung ist $\xi_v^0\neq 0$ (nicht Nullvektor), $N_0=-(\mathfrak{x}_v^0\xi_v^0)\neq 0$*). Daher gilt auch in einer gewissen Umgebung der Kurvenpunkte $\xi_v\neq 0$, $N=-(\mathfrak{x}_v\,\xi_v)\neq 0$. Wir wollen die Krümmung einer der ebenen Kurven $v=\text{const}$ untersuchen. Die Normalebene in einem ihrer Punkte durch die Kurventangente hat den Stellungsvektor $\xi_v/|\xi_v|$ [es ist ja $(\xi\xi_v)=0$ und

*) Aus $(\mathfrak{x}_v^0\,\xi_v^0)=b$ (Normalkrümmung) $=0$ würde nach $a^2+b^2+2Hb+K=0$ (GROTEMEYER) $a=0$ und nach $\xi_v^0=b\,\mathfrak{x}_v-a\eta$ (BLASCHKE, Differentialgeometrie S. 68) $\xi_v^0=0$ folgen. Hier ist v als Bogenlänge der Kurve $u=0$ angenommen.

$(\mathfrak{x}_u \xi_v) = -M = 0]$, und der zugehörige Normalschnitt hat die Krümmung $-\frac{L}{E}$, die wir auch $-\frac{LN}{EN}$ schreiben können wegen $N \neq 0$. ε sei der kleine Winkel, den die Normalebene mit der Schnittebene der Kurve $v = \text{const}$ bildet: $\cos\varepsilon = \frac{(\xi_v \xi_v^0)}{|\xi_v|\,|\xi_v^0|}$. Dann ist die Krümmung der Kurve $v = \text{const}$ in dem betrachteten Punkt $-\frac{1}{\cos\varepsilon}\frac{LN}{EN}$. Der Zähler hat wegen $M = 0$ dasselbe Vorzeichen wie die Flächenkrümmung, er wechselt beim Übergang aus dem einen Gebiet in das andere das Zeichen; auf der parabolischen Kurve verschwindet er. $\cos\varepsilon$ hat einen Wert, der beliebig wenig von 1 verschieden ist. Also hat die Kurve $v = \text{const}$ in den beiden Gebieten verschiedenes Vorzeichen der Krümmung und für $u = 0$ einen Wendepunkt. Die Tangenten in den Wendepunkten der Kurven $v = \text{const}$ sind die Erzeugenden der abwickelbaren Fläche längs der parabolischen Kurve, der „Streifenfläche". Diese durchsetzt also die gegebene Fläche, wie im Hilfssatz angegeben ist. Wir können nun behaupten:

Satz 3. *Eine Rinne läßt sich überhaupt nicht stetig verbiegen.*

Wegen Satz 1 brauchen nur Verbiegungen in Nichtrinnen diskutiert zu werden. Wollte man aber eine Nichtrinne in eine Rinne verbiegen, so müßte die abwickelbare „Streifenfläche" sich in die Ebene der parabolischen Kurve der Rinne abwickeln. Bei diesem Vorgang aber würden die beiden verschieden gekrümmten Teile der Fläche auf verschiedenen Seiten der abwickelbaren Fläche bleiben und würden also in Flächenstücke auf verschiedenen Seiten der Ebene übergehen. Also kann im Endergebnis keine Rinne resultieren, die ja ganz auf der einen Seite der Ebene der parabolischen Kurve liegen muß.

4. Während eine stetige Verbiegung der Rinne also nicht möglich ist, gilt

Satz 4. *Im Kleinen gibt es zu jeder Rinne isometrische Nichtrinnen.*

Man braucht nur den „parabolischen Streifen" von R beliebig zu verbiegen, doch so, daß er nirgends Schmiegstreifen wird, und hat dann zur Bestimmung einer durch ihn gehenden isometrischen Fläche ein reguläres Cauchysches Problem für die eine Koordinate der gesuchten Biegungsfläche zu lösen, die der Bourschen Gleichung genügen muß*). Die Bestimmung der anderen Koordinaten ist dann äquivalent dem Problem, eine quadratische Differentialform der Krümmung 0 auf die Normalform $dx^2 + dy^2$ zu bringen.

*) Man kann die Gleichung in der Form anwenden [H. Liebmann, Münchener Berichte 1920, S. 33, Gl. (13)]

$$(1 + p^2 + q^2)(\varrho\tau - \sigma^2) - (p\pi + qx)(r\tau - 2s\sigma + t\varrho) + (rt - s^2)(-1 + \pi^2 + x^2) = 0.$$

$x, y, z(x, y), p, q, r, s, t$ beziehen sich auf die Rinne mit $p = q = rt - s^2 = 0$ auf der Kurve, $\xi(x, y)$, $\eta(x, y)$, $\zeta(x, y)$, $\zeta_x = \pi$, $\zeta_y = x$, ϱ, σ, τ auf die Nichtrinne mit $\pi' = \varrho x' + \sigma y'$, $x' = \sigma x' + \tau y'$, $\varrho\tau - \sigma^2 = 0$, $\pi' x' + x' y' \neq 0$ auf der Kurve. Man erhält $\varrho : \sigma : \tau : 1 = \pi'^2 : \pi' x' : x'^2 : (\pi' x' + x' y')$. Vgl. auch G. Darboux, Surfaces **3**, livre VII, chap. V; Bianchi-Lukat, Differentialgeometrie 2. Aufl. 1910, Kap. 7, §§ 109, 110; Hoheisel, Partielle Differentialgleichungen 1928, Kap. VI, § 2.

Wir betrachten ein einfaches Beispiel, in dem man eine isometrische Nichtrinne ohne Integration der betreffenden Differentialgleichung angeben kann. Für $|u| \leqq \varepsilon$ ist

$$\mathfrak{x} = \big((1+\sin u)\cos v,\ (1+\sin u)\sin v,\ 1-\cos u\big)$$

eine Rinne, die isometrisch ist zu

$$\mathfrak{y} = \left(k(1+\sin u)\cos\frac{v}{k},\ k(1+\sin u)\sin\frac{v}{k},\ \int_0^u \sqrt{1-k^2\cos^2 u}\,du\right)$$

mit $k = \text{const} < 1$. Beide Flächen haben das Linienelement

$$du^2 + (1+\sin u)^2\,dv^2.$$

$\mathfrak{y}$ ist aus $\mathfrak{x}$ nach demselben Prinzip gebildet wie die Mindingsche Fläche konstanter positiver Krümmung mit Kegelspitzen aus der Kugel. Die Einhüllende der Tangentialebenen längs der parabolischen Kurve von $\mathfrak{y}$ ist ein Rotationskegel. Bei der stetigen Verbiegung ($k \to 1$), bei der der Kegel in die 1, 2-Ebene übergeht, geht nur der positiv gekrümmte Teil von $\mathfrak{y}$ in den von $\mathfrak{x}$, der mit negativer Krümmung dagegen in einen zum entsprechenden von $\mathfrak{x}$ spiegelbildlichen über, wie nach **3** zu erwarten war. In der Grenze entsteht eine Fläche, die auf der parabolischen Kurve nur Glattheit erster Ordnung hat.

5. Ist R insbesondere eine geschlossene Rinne, so gelten die Bemerkungen von **4** nicht mehr allgemein.

SATZ 5. *Es gibt zu einer geschlossenen Rinne keine nichtkongruente Fläche mit ebener konvexer und geschlossener parabolischer Kurve.*

Da die Tangentialebenen auf einer solchen isometrischen Kurve wechseln, müßten sie dort mit der Kurvenebene einen Winkel $\varphi \not\equiv 0$ bilden. Nun haben die Rinne und die isometrische Fläche in entsprechenden Punkten der parabolischen Kurven wegen der Isometrie gleiche geodätische Krümmung c. Bei der Rinne ist diese mit der gewöhnlichen Krümmung k der parabolischen Kurve identisch: $c = k$. Auf der isometrischen Fläche ist aber nach MEUSNIER $c = k_1 \cos\varphi$. Also ist $k_1 \geqq k$ und wenigstens auf einem Bogen der Kurven $k_1 > k$. Es müßte, da nach FENCHEL für ebene Kurven $\oint k\,ds = 2\pi$ ist, $\oint k_1\,ds > 2\pi$ sein, was (wieder nach FENCHEL) für ebene geschlossene konvexe Kurven ausgeschlossen ist.

Die Frage, ob es zu einer konvexen geschlossenen Rinne isometrische Flächen gibt, bei denen der parabolischen Kurve der Rinne eine geschlossene Raumkurve entspricht, ist ungeklärt.

Korollare. 1. Zum Satz von BOUR (1862). *Es gibt kein Stück einer Schraubenfläche mit gemischter Gaußscher Krümmung, das als Ganzes stetig auf eine Torusfläche verbogen werden kann.* 2. *Beide Arten von Halbrinnen können wohl*

zugebogen, aber nicht aufgebogen werden. 3. Auch die (nicht geschlossene) Vollrinne wird stetig verbiegbar, wenn das Auftreten einer Kante längs der parabolischen Kurve der Biegungsfläche zugelassen wird.

Ich will noch erwähnen, daß von POSNJAK auch vieleckige geschlossene Rinnen untersucht worden sind [3]; vgl. auch [4] S. 86 und 214.

Literatur

[1] REMBS, E.: Verbiegungen höherer Ordnung und ebene Flächenrinnen. Math. Z. **36**, 110–121 (1932). — [2] ALEXANDROW, A. D.: Über eine Klasse geschlossener Flächen. Mat. Sbornik **4** (46), Nr. 1, 69–77 (1938). — [3] POSNJAK, E. G.: Unendlich kleine Verbiegungen von Rinnen. Mat. Sbornik, n. S. **32** (74), 681–692 (1953). — [4] EFIMOW, N. W.: Flächenverbiegung im Großen, mit Nachtrag von E. REMBS und K. P. GROTEMEYER. Berlin 1957.

Berlin-Friedenau, Begasstraße 7

(Eingegangen am 27. September 1958)

BAIER, O.
Math. Zeitschr. 71, 94–98 (1959)

Über die Konstruktion von Parallelen mit Hilfe des Lineals und ortsfester rechter Winkel

Herrn ROBERT SAUER zum 60. Geburtstag gewidmet

Von
OTHMAR BAIER

Bekanntlich ist es unmöglich, in der euklidischen Ebene mit dem Lineal allein durch einen vorgegebenen Punkt die Parallele zu einer gegebenen Geraden zu konstruieren. Zur Lösung dieser Aufgabe sind daher andere konstruktive Hilfsmittel erforderlich (z. B. der Gebrauch des Zirkels allein wie bei den Mohr-Mascheronischen Konstruktionen) oder neben dem Lineal noch zusätzliche (z. B. der Steinerkreis). Die folgenden Betrachtungen wollen zu einer hier einschlägigen Fragestellung einen Beitrag geben. Dabei wird mehrfach auf die hinsichtlich Inhalt und Form meisterhafte Darstellung der geometrischen Konstruktionen Bezug genommen, die H. LEBESGUE diesem Gegenstand in seiner letzten Vorlesung gegeben hat[1]).

Es seien zunächst einige bekannte, für das Verständnis notwendige Sätze vorangestellt.

1. Ist ein Paar von parallelen Geraden p und p_1 gegeben, so läßt sich durch einen beliebigen Punkt P die Parallele zu den gegebenen Geraden mit dem Lineal allein konstruieren, indem zunächst ein Punkt Q der Polaren von P in bezug auf p, p_1 konstruiert wird und dann ein Punkt P_1 der Polaren von Q. PP_1 ist die gesuchte Parallele (Fig. 1). Falls die Geraden p und p_1 nicht parallel sind und sich im Punkt S schneiden, liefert die Konstruktion einen Punkt der Verbindungsgeraden PS. H. LEBESGUE bemerkt hiezu[2]), daß die Konstruktion nicht nur gestattet, P mit einem unzugänglichen Punkt S zu verbinden, sondern daß sie außerdem mit einem beliebig kurzen Lineal durchgeführt werden kann, indem das Prinzip der Konstruktion nötigenfalls mehrmals zur Durchführung der Konstruktion selbst angewandt wird, wovon man sich leicht durch einen Blick auf Fig. 1 überzeugt.

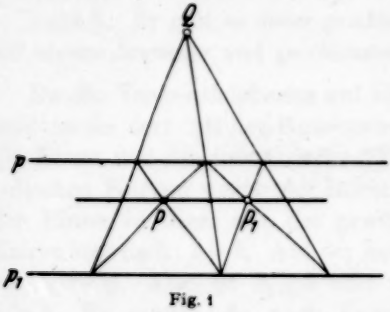


Fig. 1

2. Sind zwei Paare paralleler Geraden p, p_1 und q, q_1 gegeben, die nicht untereinander parallel sind, so kann mit dem Lineal allein durch einen beliebigen Punkt P die Parallele g_1 zu einer gegebenen Geraden g konstruiert

[1]) LEBESGUE, H.: Leçons sur les constructions géométriques professées au Collège de France en 1940/41. Paris: Gauthier-Villars 1950.

[2]) Vgl. [1]), S. 41.

werden[3]). In Fig. 2 sind die Dreiecke DP_1Q und EPQ_1 ähnlich mit dem Ähnlichkeitszentrum Z auf ED. QZ ergibt Q_1 und damit $PQ_1 = g_1 \| g$. Auch diese Konstruktion kann mit einem beliebig kurzen Lineal durchgeführt werden.

Sind in der Ebene beliebig viele rechte Winkel mit lauter verschiedenen Scheiteln gegeben, dann ist

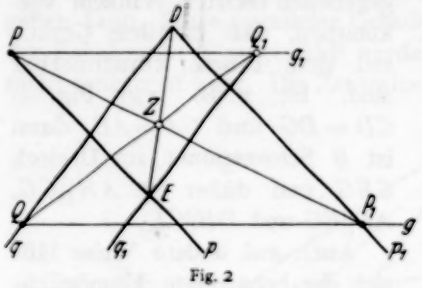


Fig. 2

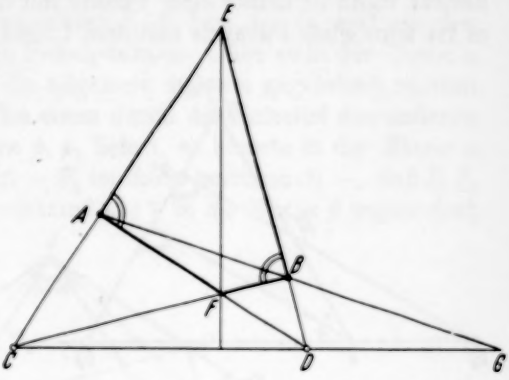


Fig. 3

es im allgemeinen unmöglich mit dem Lineal allein Paare paralleler Geraden zu konstruieren.

Der Sachverhalt möge zunächst für den Fall zweier gegebener rechter Winkel erläutert werden (vgl. Fig. 3). Die Scheitel der rechten Winkel

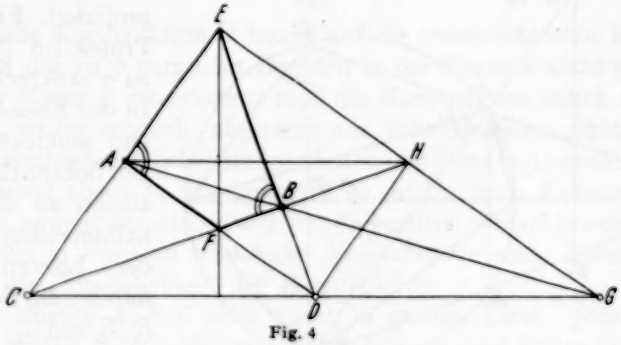


Fig. 4

befinden sich in den Punkten A und B, die Schnittpunkte der Schenkel seien die Punkte C, D, E, F. Auf Grund des Höhensatzes läßt sich noch ein weiteres Paar zueinander senkrechter Geraden, CD und EF, konstruieren, womit auf der uneigentlichen Geraden durch die drei Paare von senkrechten Geraden drei Punktepaare der absoluten Involution ausgeschnitten werden. Führt man ein projektives Koordinatensystem ein, etwa mit A, C, D als Grundpunkten und B als Einheitspunkt, dann sind mit dem Lineal allein genau die Punkte konstruierbar, deren projektive Koordinaten sich zueinander wie drei ganze Zahlen verhalten. Liegen die beiden rechten Winkel z. B. zueinander

[3]) Dies ist auf verschiedene Arten möglich, vgl. [1]), S. 42. Wegen 1. kann angenommen werden, daß p durch P geht.

so, daß die uneigentliche Gerade der Ebene die Gleichung

$$\sqrt{2}\,x_1 + \sqrt{3}\,x_2 + x_3 = 0$$

besitzt, dann ist keiner ihrer Punkte mit dem Lineal allein konstruierbar, d.h. es ist unmöglich Parallele mit dem Lineal allein zu konstruieren.

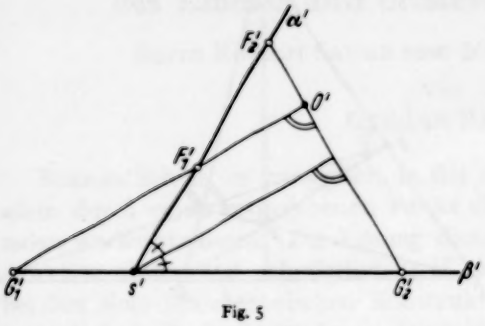


Fig. 5

Bei speziellen Abmessungen kann es natürlich bereits bei zwei gegebenen rechten Winkeln vorkommen, daß parallele Gerade mit dem Lineal konstruierbar sind. Ist z. B. (vgl. Fig. 4) $CD = DG$ und $CA = AE$, dann ist B Schwerpunkt im Dreieck CEG und daher ist $AH \| CG$, $AD \| EG$ und $DH \| CE$.

Auch auf andere Weise läßt sich die behauptete Unmöglichkeit einer allgemeinen Konstruktion von Parallelen mit dem Lineal allein bei vorgegebenen rechten Winkeln fast unmittelbar einsehen, indem man eine bekannte Eigenschaft der Zentralkollineation heranzieht, die hier elementar hergeleitet werden soll. Zwei Ebenen α und β mit der Schnittlinie s werden von O aus aufeinander projiziert. Fig. 5 ist die Projektion $\| s$ auf eine zu s senkrechte Ebene. In der Ebene γ durch O und senkrecht s schneiden bekanntlich die Parallelen zu den winkelhalbierenden Geraden der Ebenen α und β durch O die Fokalpunkte F_1, F_2 und G_1, G_2 in α und β aus. Wird die Ebene α um s in die Ebene gedreht, dann kommen F_1 und G_1 oder F_2 und G_2 (vgl. Fig. 6) zur Deckung, wobei das Kollineationszentrum in die zusammenfallenden Fokalpunkte gelangt. In den Fokalpunkten ist die Abbildung offensichtlich konform, mit bzw. ohne Änderung des Drehsinnes. Sind P und Q ein Paar einander durch die Kollineation zugeordneter Punkte, dann gibt eine bekannte Konstruktion ($PN = NQ$, $NM \perp PQ$) die zueinander senkrechten Geraden durch P, welchen wieder senkrechte Gerade durch Q entsprechen (zugeordnete Rechtwinkelstrahlen). Auf Grund des Peripheriewinkelsatzes sind die gleichmarkierten Winkel jeweils einander gleich. Daraus folgt, daß die Rechtwinkelstrahlen in P

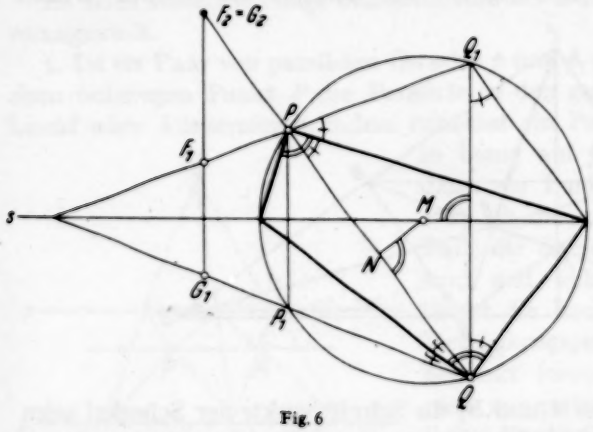


Fig. 6

und Q die Winkel der Fokalstrahlen dieser Punkte halbieren. Somit besteht die Gesamtheit der zueinander senkrechten Geraden in der Ebene α, welchen in der Ebene β wieder zueinander senkrechte Geraden entsprechen und umgekehrt, aus den Tangenten und Normalen in den Punkten der konfokalen Kegelschnitte mit den Brennpunkten F_1, F_2 bzw. G_1, G_2 und aus den rechten Winkeln mit Scheiteln in den Fokalpunkten. Gäbe es in der Ebene α eine Konstruktion mit dem Lineal, die allgemein zu zwei gegebenen rechten Winkeln, wobei auch ein Schenkel des einen durch den Scheitel des anderen gehen kann, Paare paralleler Geraden p, p_1 liefert, so könnte in der Ebene α der Fokalpunkt F_1 so gewählt werden – F_2 ist dann bestimmt[4]) –, daß $F_1 F_2$ nicht senkrecht p ist. Die Zentralprojektion aus 0 in die Ebene β ergibt dort

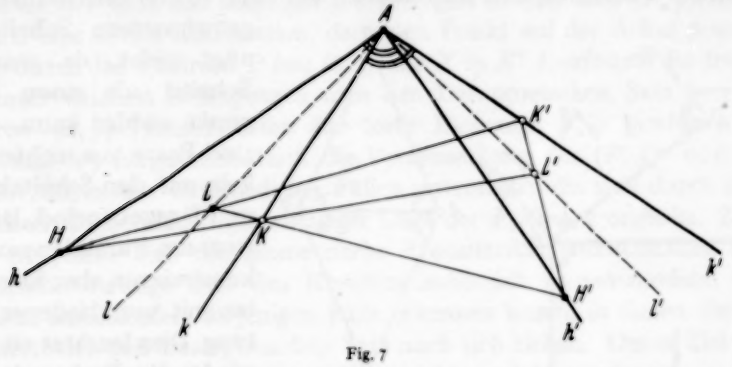


Fig. 7

wieder dieselbe Konstruktion in bezug auf die entsprechenden Rechtwinkelstrahlen. Da den zu p parallelen Geraden in der Ebene α nicht parallele Geraden in der Ebene β entsprechen, muß die Konstruktion falsch sein.

Wäre es weiter möglich, allgemein aus beliebig vielen rechten Winkeln mit lauter verschiedenen Scheiteln mindestens ein Paar von parallelen Geraden mit dem Lineal allein zu konstruieren, so müßte diese Konstruktion auch auf die Fälle anwendbar sein, in welchen die rechten Winkel jeweils Tangenten und Normalen in Punkten konfokaler Kegelschnitte sind. Dieselbe Schlußweise zeigt die Unmöglichkeit der Konstruktion.

Die drei Punkte A, B, C seien nicht in gerader Linie. Jeder von ihnen sei Scheitel zweier rechter Winkel. Dann können zwei Paare von parallelen Geraden mit dem Lineal allein konstruiert werden, sogar mit einem beliebig kurzen. Durch zwei rechte Winkel (h, h') und (k, k') mit gemeinsamem Scheitel ist in diesem die Rechtwinkelinvolution bestimmt. In jeder Strahleninvolution kann der einem beliebigem Strahl zugeordnete mit dem Lineal allein konstruiert werden. In Fig. 7 seien die beiden Paare h, h' und k, k' durch A gegeben. Gesucht ist der zu l senkrechte Strahl l' durch A. Die Punkte H auf h, H auf h', L auf l seien verschieden von A, nicht in gerader Linie, sonst aber beliebig gewählt. Es sei

$$H'L \times k = K, \quad HL \times k' = K, \quad HK \times H'K' = L .$$

[4]) Im allgemeinen durch eine Cremonatransformation, worauf hier nicht näher eingegangen wird.

Die Punkte H, H', K, K', L, L' sind die Ecken eines vollständigen Vierseits, das von A (dual zum Satz von DESARGUES über das vollständige Viereck) durch eine Strahleninvolution projiziert wird und AL' ist damit der gesuchte Strahl l'. Falls das verwendete Lineal zu kurz ist um die verlangten Geraden zu ziehen, werden zuvor durch die beliebig gewählten Punkte willkürliche Geraden gezeichnet, und dann kann man nach 1. verfahren. Damit sind (Fig. 8) in A und B die Lote zu AB sowie in B und C die Lote zu BC konstruierbar, womit zwei Paare paralleler Geraden mit dem Lineal allein konstruiert sind.

Man kann die Frage stellen, ob vielleicht weniger rechte Winkel zur Konstruktion der Parallelenpaare ausreichen. Ein Paar von rechten Winkeln mit gemeinsamem Scheitel genügt nicht, da man den Scheitel als einen Fokalpunkt wählen kann. Wenn zwei Paare von rechten Winkeln mit den Scheiteln in A und B gegeben sind, läßt sich zwar das Parallelenpaar p, p_1 konstruieren aber kein zweites mit verschiedener Richtung. Dies leuchtet ein, wenn wieder die Punkte A und B als Fokalpunkte F_1 und F_2 in der Ebene α gewählt werden.

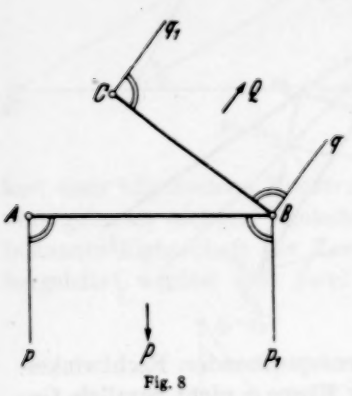


Fig. 8

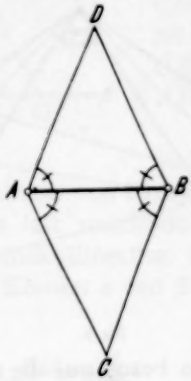


Fig. 9

Auch dann, wenn in jedem dieser Fälle noch beliebig viele rechte Winkel mit lauter verschiedenen Scheiteln dazu gegeben sind, kann es keine allgemeine Konstruktionsvorschrift geben, die zwei Paare paralleler Geraden mit verschiedener Richtung liefert. Die Vorschrift müßte sich auch auf die Fälle anwenden lassen, in welchen die rechten Winkel von den Tangenten und Normalen in Punkten konfokaler Kegelschnitte gebildet werden. Die Zentralprojektion mit den Brennpunkten als Fokalpunkten ergibt wieder einen Widerspruch.

Wie ist die Sachlage, wenn die rechten Winkel nicht stimmen, d.h. wenn kongruente, aber von rechten Winkeln verschiedene Winkel gegeben sind? Es genügen dann zwei Tripel dieser Winkel mit den Scheiteln A und B. An Stelle der Strahleninvolution tritt jeweils eine Projektivität, die durch drei Paare entsprechender Strahlen festgelegt ist. Mit dem Lineal allein läßt sich zu jedem Strahl der entsprechende konstruieren. Somit ist die Raute $ACBD$ (Fig. 9) konstruierbar. Mit Hilfe ihrer parallelen Gegenseiten kann zu jeder Geraden die Parallele durch einen beliebigen Punkt mit dem Lineal allein konstruiert werden und wieder mit einem beliebig kurzen.

Stuttgart, Mathematisches Institut B der Technischen Hochschule

(Eingegangen am 4. Oktober 1958)

PICKERT, G.
Math. Zeitschr. 71, 99–108 (1959)

Eine Kennzeichnung desarguesscher Ebenen

Von
GÜNTER PICKERT

Bekanntlich ist eine desarguessche projektive Ebene für jeden Punkt P und jede Gerade p (P, p)-transitiv, d.h.: Zu Punkten X, X', die mit P kollinear aber $\neq P$ sind und nicht auf p liegen, gibt es stets eine (P, p)-Kollineation (d.i. eine solche Kollineation, die jeden Punkt auf der Achse p und jede Gerade durch das Zentrum P fest läßt), die X in X' überführt. Es fragt sich nun, unter welchen Bedingungen man den Desarguesschen Satz bereits aus mehreren (P, p)-Transitivitäten für feste Elemente P, p gewinnen kann. T. G. OSTROM[1]) hat systematisch die Voraussetzung der (P, p)- und (Q, q)-Transitivität in den verschiedenen Fällen untersucht, die sich durch die verschiedenen Möglichkeiten gegenseitiger Lage der P, Q, p, q ergeben. Ziel war dabei, in jedem Fall die geometrische Transitivitätsvoraussetzung in eine algebraische Aussage über den Koordinatenbereich zu verwandeln, woraus sich dann insbesondere diejenigen Fälle erkennen lassen, in denen die beiden Transitivitäten den Desarguesschen Satz nach sich ziehen. Dieses Ziel konnte lediglich in den folgenden beiden Fällen nicht erreicht werden:

(1) $\qquad P \neq Q, \quad p \neq q; \quad P, Q \notin p, q; \quad p \cap q \in PQ.$

(2) $\qquad P \neq Q, \quad p \neq q; \quad P \in p, Q \in q; \quad P \notin q, \quad Q \notin p.$

Für Fall (2) zeigen Beispiele, daß hier nicht der Desarguessche Satz, ja noch nicht einmal der kleine Desarguessche Satz zu gelten braucht. Für Fall (1) dagegen sind solche Beispiele nicht bekannt.

Im folgenden zeige ich in Satz 2, daß sich im Fall (1) der Desarguessche Satz ergibt, wenn noch die $(P, Q, p \cap q)$-*Thomsen-Bedingung* hinzugenommen wird, d.i. die Thomsen-Bedingung[2]) im Gewebe der drei Büschel mit den Trägern $P, Q, p \cap q$. Da die Thomsen-Bedingung als nicht durch Transitivitätsaussagen ersetzbar hier etwas aus dem Rahmen fällt, wäre es natürlich schön, wenn man ohne sie auskommen würde. Immerhin läßt sich ein beim Beweis von Satz 2 benötigtes Teilergebnis bereits ohne die Thomsen-Bedingung herleiten (Satz 1). Da der Koordinatenbereich einer desarguesschen Ebene ein Schiefkörper ist, liefert Satz 2 nebenbei noch eine merkwürdige Kennzeichnung der Schiefkörper: Statt der Distributivgesetze wird gefordert, daß die vom Einselement verschiedenen Elemente eine Gruppe bezüglich der Verknüpfung $(a, b) \to a - ab + b$ bilden (Satz 3).

[1]) Transitivities in projective planes. Can. J. of Math. 9, 389–399 (1957).

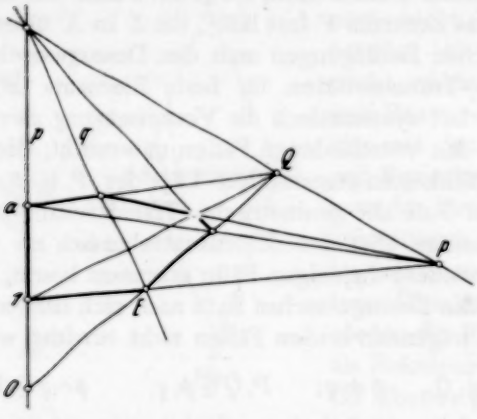
[2]) Siehe etwa G. PICKERT: Projektive Ebenen. Berlin-Göttingen-Heidelberg 1955 (im folgenden als *PE* zitiert); S. 59.

Satz 1. *Genügen die Punkte P, Q und die Geraden p, q einer projektiven Ebene der Bedingung* (1) *und ist die Ebene (P, p)- sowie (Q, q)-transitiv, so gelten die folgenden Aussagen:*

(a) *Es gibt eine $(p \cap q, PQ)$-Kollineation, die für jeden Punkt R von q den Punkt $p \cap RQ$ in den Punkt $p \cap RP$ überführt.*

(b) *Es gilt die $(P, Q, p \cap q)$-Thomsen-Bedingung mit der Einschränkung, daß in ihr zwei der durch $p \cap q$ gehenden Geraden mit den Geraden p, q zusammenfallen (s. Figur).*

(c) *Ist die Ebene nicht desarguessch, so gilt in jedem Koordinatensystem mit $P, p \cap q$ als den uneigentlichen Punkten U, V der Koordinatenachsen, mit*



Koordinatenanfangspunkt O auf p und mit $E = q \cap OQ$ als dem Einheitspunkt[3]) *die Beziehung $a + (a^{-1} + b) = 1 + b$ für alle Koordinaten $a \neq 0$ und $\neq 1$ und alle Koordinaten b.*

Beweis. Da (a) und (b) aus dem Satz von Desargues folgen, beweise ich zuerst (c) und dann daraus im nichtdesarguesschen Fall (a) und (b). Die Bestimmungsstücke U, V, O, E eines Koordinatensystems seien wie in (c) gewählt, also $U = P, V = p \cap q, O \in p, E = q \cap OQ$. Wegen der (P, p)-Transitivität läßt sich jede nicht durch V gehende Gerade durch eine Gleichung $y = ux + v$ für die Koordinatenpaare (x, y) ihrer Punkte darstellen, die Multiplikation ist assoziativ und für jede Koordinate $a \neq 0$ ist

$$(x, y) \to (ax, y) \tag{3}$$

die Koordinatendarstellung einer (P, p)-Kollineation (*PE*, S. 102–103, Satz 45). U_c bezeichne den Schnittpunkt der durch $y = cx$ dargestellten Geraden mit UV und O_x den Punkt mit dem Koordinatenpaar $(x, 0)$, so daß also $Q = U_1$, $P = U_0$, $0 = O_0$ ist. Die durch (3) dargestellte Kollineation führt dann offenbar O_1 in O_a, also q in VO_a, und Q in $U_{a^{-1}}$ über, so daß die Ebene

[3]) Koordinatenpaar $(1, 1)$; zu den Bezeichnungen vgl. *PE*, S. 34–35.

daher auch $(U_{a^{-1}}, VO_a)$-transitiv ist. Zu jedem Punkt $Z \in UV$, $\neq V$ gibt es also eine Gerade z durch V so, daß die Ebene (Z, z)-transitiv ist. Die Gerade z ist durch den Punkt Z eindeutig bestimmt[4]). Andernfalls wäre nämlich die Ebene sogar (Z, V)-transitiv [d.h. (Z, z')-transitiv für jede Gerade z' durch V], woraus sich mittels einer (P, p)- oder (Q, q)-Kollineation die (Z', V)-Transitivität und damit die (V, Z')-Transitivität (PE, S. 103, Satz 46) für einen Punkt $Z' \neq V, Z$ von UV ergeben würde; aus (Z, V)- und (V, Z')-Transitivität zusammen folgt aber der Desarguessche Satz (PE, S. 105, Satz 49), der hier ja gerade nicht gelten sollte. Da die durch $y = a^{-1}x$ dargestellte Gerade $OU_{a^{-1}}$ die Gerade VO_a in dem auf EU gelegenen Punkt mit dem Koordinatenpaar $(a, 1)$ trifft, gilt demnach

$$z = (OZ \cap EU)\, V \quad \text{für} \quad Z \neq U. \tag{4}$$

Diese Gleichung muß nun auch bestehen bleiben, wenn bei der Wahl des Koordinatensystems die Paare P, p und Q, q ihre Rollen tauschen; denn die Voraussetzungen von Satz 1 sind ja in diesen beiden Paaren symmetrisch. Ein dieser Vertauschung entsprechendes Koordinatensystem bekomme ich nun, wenn ich $U' = Q$, $V' = V$ setze und für O', E' die Punkte mit den Koordinatenpaaren $(1, c)$, $(0, c)$ nehme, wobei c völlig beliebig sein darf. Nach (4) gehen im Falle $Z \neq U'$ dann die Geraden $O'Z$, $E'U'$ und z durch einen Punkt. $z = VO_a$ hat die Gleichung $x = a$ und $E'U'$ $(= E'U_1)$ die Gleichung $y = x + c$. Wegen $Z = U_{a^{-1}}$ erhalte ich für $O'Z$ eine Gleichung der Gestalt $y = a^{-1}x + b$, in der wegen $O' \in O'Z$

$$c = a^{-1} + b \tag{5}$$

sein muß. Die von den drei Geraden im Falle $Z \neq U'$ $(= U_1)$, also $a \neq 1$, zu erfüllende Bedingung ergibt

$$a + c = 1 + b. \tag{6}$$

Zusammen besagen (5) und (6) nichts weiter als

$$a + (a^{-1} + b) = 1 + b \quad \text{für} \quad a \neq 0, 1; \tag{7}$$

denn durch (5) wird ja eine umkehrbare Abbildung $b \to c$ des gesamten Koordinatenbereichs auf sich geliefert, so daß statt c auch b beliebig vorgegeben werden kann. Damit ist die Behauptung (c) bewiesen.

Für spätere Verwendung führe ich noch den Sonderfall $b = 0$ von (7) auf:

$$a + a^{-1} = 1 \quad \text{für} \quad a \neq 0, 1. \tag{8}$$

Hieraus ergibt sich übrigens, daß eine topologische Ebene, deren Topologie weder die feinste noch die gröbste ist (insbesondere also auch eine angeordnete

[4]) Das ergibt sich auch sofort aus der Klasseneinteilung von A. Barlotti, Le possibili configurazioni del sistema delle coppie punto-retta (A, a) per cui un piano grafico risulta (A, a)-transitivo. Boll. U.M.I. **12**, 212–226 (1957).

Ebene), unter den Voraussetzungen von Satz 1 desarguessch sein muß: Läßt man in (8) a gegen 1 streben, so ergibt sich ein Widerspruch.

Ersetzen von b durch $a+b$ in (7) liefert

$$a+(a^{-1}+(a+b))=1+(a+b)$$

und ersetzen von a durch a^{-1} wegen $(a^{-1})^{-1}=a$

$$a^{-1}+(a+b)=1+b.$$

Zusammen erhalte ich also

$$a+(1+b)=1+(a+b), \tag{9}$$

insbesondere für $b=0$

$$a+1=1+a. \tag{10}$$

Bei (9) und damit bei (10) darf offensichtlich die Einschränkung $a\neq 0, 1$ weggelassen werden. Mittels (10) wird nun (9) zu

$$a+(b+1)=(a+b)+1. \tag{11}$$

Wegen der Assoziativitätsaussage (11) ist

$$(x, y)\to(x, y+1) \tag{12}$$

die Koordinatendarstellung einer Kollineation, und zwar einer (V, UV)-Kollineation (PE, S. 100). Für den Punkt S auf p mit dem Koordinatenpaar $(0, s)$ hat SQ die Gleichung $y=x+s$ und daher $R=SQ\cap q$ das Koordinatenpaar $(1, 1+s)$. Wegen (10) ist also $p\cap RP$ der Bildpunkt von $S=p\cap RQ$ bei der durch (12) dargestellten Kollineation. Damit habe ich (a) bewiesen. (b) aber ist weiter nichts als die geometrische Übersetzung von (10), wenn man das Koordinatensystem so mit der Thomsen-Figur koppelt wie in der Figur auf S. 100 (vgl. PE, S. 59), worin die die beiden Punkte auf p mit den y-Koordinaten a und 1 einfach durch diese Koordinaten bezeichnet worden sind. Zu beachten ist, daß in der Thomsen-Bedingung von den drei Geraden durch einen der Trägerpunkte (hier handelt es sich nur um $p\cap q$) keine ausgezeichnet ist; man erkennt das leicht daraus, daß durch die Vertauschung der mit a und 1 bezeichneten Punkte zwei der Geraden durch $p\cap q$ vertauscht werden, während (10) in sich übergeht[5]).

Satz 2. *Eine projektive Ebene ist desarguessch, wenn es in ihr Punkte P, Q und Geraden p, q mit* (1) *so gibt, daß die $(P, Q, p\cap q)$-Thomsen-Bedingung erfüllt und die Ebene sowohl (P, p)- wie (Q, q)-transitiv ist.*

Beweis. Das Koordinatensystem sei wie beim Beweis von Satz 1 gewählt. Wie schon erwähnt, besagt die (P, p)-Transitivität dann, daß jede nicht durch V gehende Gerade eine Gleichung der Form $y=ux+v$ hat und daß die Multiplikation assoziativ ist. Die $(P, Q, p\cap q)$-Thomsen-Bedingung nun ist gleich-

[5]) Statt dessen kann man natürlich auch eine gewisse Permutation der Eckpunkte des Sechsecks (in der Figur auf S. 100 stärker ausgezogen) vornehmen.

wertig mit der Assoziativität und Kommutativität der Addition (*PE*, S. 52, 59). Der Koordinatenbereich ist also eine kommutative cartesische Gruppe (*PE*, S. 90) mit assoziativer Multiplikation. Ich übersetze jetzt die (Q, q)-Transitivität ebenfalls in eine algebraische Aussage. Die (Q, q)-Transitivität besagt, daß zu jedem Koordinatenwert $a \neq 1$ eine (Q, q)-Kollineation existiert, die O in den Punkt mit dem Koordinatenpaar (a, a) überführt. Da bei einer solchen Punkt und Bildpunkt kollinear mit Q $(= U_1)$ sind, folgt für die Koordinatendarstellung $(x, y) \to (x', y')$ der Kollineation

$$x' - x = y' - y. \tag{13}$$

Da jede Gerade durch V wieder in eine solche übergehen muß, hängt x' nicht von y ab. Es gibt also eine Funktion f mit

$$x' = f(x) + x, \tag{14}$$

und aus (13) folgt dann

$$y' = f(x) + y. \tag{15}$$

Um jetzt f zu bestimmen, betrachte ich das Urbild der Geraden OU. Werden die Punkte mit den Koordinatenpaaren (a, a), $(1, a)$ durch A bzw. A' bezeichnet, so ist A' Fixpunkt, und OA' geht daher in AA', also $OA' \cap UV$ in $AA' \cap UV = U$ über. Da OA' aber die Gleichung $y = a\,x$ hat, ist $OA' \cap UV = U_a$. Daher hat das Urbild von OU eine Gleichung $y = a\,x + b$. Wegen $y' = 0$ für die Bildpunkte ergibt sich daraus mit (14), (15)

$$0 = f(x) + a\,x + b \tag{16}$$

für alle x mit einem von x unabhängigen Wert b. Da O in A übergeht, liefert (14) noch $f(0) = a$, so daß nach (16) (mit $x = 0$) $b = -a$ ist. Damit wird (16) zu $f(x) = a - a\,x$, worauf (14), (15) in

$$x' = a - a\,x + x, \qquad y' = a - a\,x + y \tag{17}$$

übergehen. Da offenbar (17) für $a \neq 1$ eine umkehrbare Abbildung der Menge aller Punkte $\notin UV$ auf sich darstellt, die alle Punkte auf q festläßt, besagt die (Q, q)-Transitivität: Für $a \neq 1$ geht durch (17) jede Gerade wieder in eine Gerade über und insbesondere jede Gerade durch U_1 in eine ebensolche. Hat die Gerade die Gleichung $x = c$, so folgt aus (17) $x' = a - a\,c + c$ für die Bildgerade. Daher kann ich mich bei der folgenden algebraischen Umformung auf Geraden mit einer Gleichung $y = u\,x + v$ beschränken, die also bestimmt nicht eine Bildgeradengleichung $x' = c'$ besitzen. Daß die Bildpunkte auf einer Geraden der Gleichung $y' = u'\,x + v'$ liegen, besagt nach (17)

$$a - a\,x + u\,x + v = u'(a - a\,x + x) + v' \tag{18}$$

für alle Koordinaten x. Insbesondere folgt daraus für $x = 0$

$$a + v = u'\,a + v',$$

so daß (18) zu

$$a - ax + ux = u'(a - ax + x) - u'a + a \tag{19}$$

wird. Mit $x = 1$ erhalte ich daraus

$$u = u' - u'a + a \tag{20}$$

und, dies in (19) eingesetzt, weiter

$$a - ax + (u' - u'a + a)\,x = u'(a - ax + x) - u'a + a. \tag{21}$$

Da nun (20) eine von v unabhängige umkehrbare Abbildung $u \to u'$ des Koordinatenbereichs auf sich beschreibt, bei der 1 festbleibt, ist nach dem bereits Festgestellten die Gültigkeit von (21) für alle u', x und alle $a \neq 1$ gleichwertig mit der (Q, q)-Transitivität. Da (21) offenbar für $a = 1$ sowieso gilt, kann dabei die Einschränkung $a \neq 1$ noch weggelassen werden. Um die Bedingung (21) übersichtlicher schreiben zu können, führe ich durch

$$a \circ b = a - ab + b \tag{22}$$

im Koordinatenbereich die neue Verknüpfung $\circ$ ein, die im folgenden bezüglich Klammersetzung in Verbindung mit der Addition genau so wie die Multiplikation behandelt sei. (21) wird nun der Reihe nach zu:

$$\begin{gathered} a \circ x - x + (u' \circ a)\,x = u'(a \circ x) - u' + u' \circ a, \\ -(u' \circ a)\,x + x - a \circ x = -u' \circ a + u' - u'(a \circ x), \\ u' \circ a - (u' \circ a)\,x + x = u' - u'(a \circ x) + a \circ x, \\ (u' \circ a) \circ x = u' \circ (a \circ x). \end{gathered}$$

Damit habe ich das Zwischenergebnis: *Die (Q, q)-Transitivität bedeutet die Assoziativität der in (22) erklärten Verknüpfung $\circ$.*

Daß diese Assoziativität eine Folge der (Q, q)-Transitivität ist, läßt sich übrigens ohne Rechnung leicht folgendermaßen aus (17) gewinnen. Hinter der durch (17) dargestellten Kollineation denke ich mir diejenige ausgeführt, die durch (17) mit b statt a dargestellt wird. So entsteht wieder eine (Q, q)-Kollineation; diese sei durch (17) mit c statt a dargestellt. Die jeweils ersten Gleichungen in den Koordinatendarstellungen der drei Kollineationen ergeben dann für alle x die Gleichung $b \circ (a \circ x) = c \circ x$. Mit $x = 0$ folgt daraus $b \circ a = c$ und damit für alle x die Gleichung $b \circ (a \circ x) = (b \circ a) \circ x$, die ja in den bei dieser Herleitung ausgeschlossenen Fällen $a = 1$ und $b = 1$ sowieso gilt. Da ich aber für den Beweis von Satz 3 auch benötige, daß die Assoziativität für die (Q, q)-Transitivität hinreicht, habe ich hier den obigen längeren Beweis gegeben, der dies gleich mit liefert.

Im folgenden sei jetzt die Ebene als nichtdesarguessch angenommen; daraus leite ich dann einen Widerspruch her. Die nach Satz 1 (c) zur Verfügung

stehende Gl. (8) schreibt sich wegen (22) als

$$a^{-1} \circ a = 0 = a \circ a^{-1} \quad \text{für} \quad a \neq 0, 1.$$

Wegen der Assoziativität von $\circ$ folgt daraus

$$a^{-1} \circ (a \circ b) = (a^{-1} \circ a) \circ b = 0 \circ b = b,$$

also

$$a^{-1} - a^{-1}(a \circ b) + a - ab + b = b,$$

$$-a^{-1}(a \circ b) = -a^{-1} + ab - a,$$

$$a \circ b = a(a - ab + a^{-1}).$$

Nach (8) und wegen der Kommutativität der Addition ist nun

$$a - ab + a^{-1} = (a + a^{-1}) - ab = 1 - ab = (ab)^{-1} = b^{-1}a^{-1},$$

wenn noch $ab \neq 0, 1$ vorausgesetzt wird. Daher ergibt sich

(23) $$a \circ b = ab^{-1}a^{-1}, \quad \textit{falls} \quad a \neq 0, 1 \quad \textit{und} \quad b \neq 0, a^{-1}.$$

Von jetzt ab betrachte ich nur noch die multiplikative Gruppe des Koordinatenbereichs; d. h. alle vorkommenden Koordinaten sollen $\neq 0$ sein. In dieser Gruppe kommutiert ein Element $a \neq 1$ nur mit 1, a und a^{-1}; denn aus $ab = ba$ und $b \neq a^{-1}$, 1 folgt nach (23) wegen der Kommutativität der Addition:

$$b^{-1} = a \circ b = a + b - ab = b + a - ba = b \circ a = a^{-1},$$

also $b = a$. Da eine Ebene mit weniger als 9 Punkten auf einer Geraden desarguessch ist[6]), hier aber eine nichtdesarguessche Ebene vorausgesetzt wurde, enthält die multiplikative Gruppe mindestens 7 Elemente. Sie kann also nach dem eben über kommutierende Elemente Festgestellten insbesondere nicht kommutativ sein, und es gibt daher in ihr ein Element a mit $a^2 \neq 1$. Da a^2 mit a kommutiert, wegen $a \neq 1$ aber $a^2 \neq a$ gilt, muß dann

(24) $$a^3 = 1$$

gelten. Es gibt nun Gruppenelemente $\neq 1, a, a^{-1}$, also solche, die nicht mit a kommutieren. Deren Quadrate sind nicht sämtlich $= 1$. Andernfalls könnte man nämlich so schließen: Zu $b \neq 1, a, a^{-1}$ gibt es aus Anzahlgründen ein von den $1, a, a^{-1}, b, ba, ba^{-1}$ verschiedenes c, und wegen $b^2 = 1$ ist $bc \neq 1, a, a^{-1}$, also $(bc)^2 = 1$; daraus folgt nun $bc = (bc)^{-1} = c^{-1}b^{-1} = cb$ im Widerspruch dazu, daß b, c wegen $1 \neq b = b^{-1} \neq c \neq 1$ nicht miteinander kommutieren können. Es gibt also ein nicht mit a kommutierendes Element b, für das $b^2 \neq 1$ gilt.

[6]) Statt 9 kann man hier auch 10 setzen (s. das Ergebnis von M. Hall, J. D. Swift u. R. J. Walker in: Math. Tables and Other Aids to Computation **10**, 186–194 (1956).

Nach (23) ergibt sich dann:

$$a \circ b = ab^{-1}a^{-1} \quad \text{wegen} \quad a \neq 1, b \neq a^{-1};$$
$$(a \circ b) \circ b = ab^{-1}a^{-1}b^{-1}aba^{-1} \quad \text{wegen} \quad ab^{-1}a^{-1} \neq 1, \quad b^{-1} \neq ab^{-1}a^{-1};$$
$$b \circ b = b^{-1} \quad \text{wegen} \quad b \neq 1, \quad b^{-1} \neq b;$$
$$a \circ (b \circ b) = aba^{-1} \quad \text{wegen} \quad a \neq 1, b^{-1} \neq a^{-1}.$$

Auf Grund der Assoziativität von $\circ$ ist also $ab^{-1}a^{-1}b^{-1} = 1$, d.h.

$$ba = ab^{-1}. \tag{25}$$

Nun kommutieren aber auch b^{-1}, a nicht miteinander, und es ist $a^2 \neq 1$; daher gilt neben (25) noch $ab^{-1} = b^{-1}a^{-1}$ und daher $ba = b^{-1}a^{-1}$, d.h.

$$b^2a^2 = 1. \tag{26}$$

In dieser Gleichung darf man natürlich b durch b^{-1} ersetzen, da ihre einzige Voraussetzung neben $b^2 \neq 1$ das Nichtkommutieren von a, b war. So ergibt sich

$$b^2 = a^2. \tag{27}$$

Aus den Gln. (24), (26), (27) erhalte ich schließlich $a = 1$ im Widerspruch zur Voraussetzung.

Satz 3. *Eine Menge mit einem Paar von binären, als Addition und Multiplikation geschriebenen Verknüpfungen, die in der Menge überall erklärt sind, ist genau dann ein Schiefkörper, wenn folgende Bedingungen erfüllt sind:*

(a) *Die Menge ist mit der Addition eine kommutative Gruppe.*

(b) *Die Menge ohne das neutrale Element* 0 *der additiven Gruppe ist mit der Multiplikation eine Gruppe, und für deren neutrales Element* 1 *gilt*

$$0 \cdot 1 = 0 = 1 \cdot 0.$$

(c) *Die Menge ohne* 1 *ist mit der durch* $a \circ b = a - ab + b$ *erklärten Verknüpfung* $\circ$ *eine Gruppe.*

Beweis. Daß ein Schiefkörper nicht nur (a), (b) sondern auch (c) erfüllt, erkennt man am einfachsten so: Da sich die Definition von $\circ$ in einem Schiefkörper zu $1 - a \circ b = (1 - a)(1 - b)$ umformen läßt, ist $a \to 1 - a$ ein Isomorphismus der Menge der Schiefkörperelemente $\neq 1$ mit der Verknüpfung $\circ$ auf die multiplikative Gruppe. Um nun zu zeigen, daß die Bedingungen (a), (b), (c) auch ausreichen, leite ich aus ihnen zuerst her, daß die Menge mit Addition und Multiplikation eine kommutative cartesische Gruppe mit assoziativer Multiplikation ist. Nach (b) gilt $1 \neq 0$. Wird das neutrale Element der in (c) genannten Gruppe mit e bezeichnet, so gilt wegen der Definition von $\circ$ und der in (b) enthaltenen Gleichung für alle Elemente a

$$ae = e = ea. \tag{28}$$

Insbesondere ist $e^2 = e$, also $e = 0$ wegen $e \neq 1$. Daher liefert (28) die Aussage

$$a0 = 0 = 0a. \tag{29}$$

Da die Addition nach (a) kommutativ sowie assoziativ und die Multiplikation nach (b) und (29) assoziativ ist, brauche ich nur noch zu zeigen (s. *PE*, S. 90), daß die Gleichung

$$-ax + bx = c \tag{30}$$

ebenso wie die Gleichung

$$xa - xb = c \tag{31}$$

bei gegebenen a, b, c mit $a \neq b$ durch genau je ein Element x erfüllt wird. Der Fall $b = 0$ bei (30) und der Fall $a = 0$ bei (31) erledigen sich nach (b) und (29). Bei $b \neq 0$ mit $y = bx$ wird (30) nun zu

$$(ab^{-1}) \circ y = ab^{-1} + c$$

und (31) bei $a \neq 0$ mit $y = xa$ zu

$$y \circ (a^{-1}b) = c + a^{-1}b;$$

dabei ist $ab^{-1} \neq 1$ und $a^{-1}b \neq 1$. Die eindeutige Auflösbarkeit dieser Gleichungen — und damit auch die der ursprünglichen wegen (b) und (29) — folgt nun sofort aus (c), wenn man noch bedenkt, daß $z \circ 1 = 1 = 1 \circ z$ für alle z und (wegen der Gruppeneigenschaft der Menge der Elemente $\neq 1$ bezüglich $\circ$) $z \circ z' \neq 1$ im Falle $z, z' \neq 1$ ist.

Die cartesische Gruppe läßt sich nun als Koordinatenbereich einer projektiven Ebene auffassen, die [mit der in Satz 1(c) erklärten Bedeutung der U, V, O, E) (U, OV)-transitiv ist und der $(U, V, OE \cap UV)$-Thomson-Bedingung genügt. Nach dem Beweis von Satz 2 besagt die Assoziativität der Verknüpfung $\circ$ dann die $(OE \cap UV, EV)$-Transitivität. Aus Satz 2 folgt nun der Desarguessche Satz und damit die Schiefkörpereigenschaft des Koordinatenbereichs. Damit ist Satz 3 bewiesen.

Das im Anschluß an Gl. (8) Bemerkte zeigt noch, daß bei Vorhandensein einer nichtdiskreten T_2-Topologie, bezüglich der Addition und Inversenbildung stetig sind, die Kommutativität der Addition in (a) weggelassen werden kann. Wenn dies auch allgemein (d. h. ohne die topologischen Zusatzvoraussetzungen) richtig wäre, würde das für Satz 2 besagen, daß die $(P, Q, p \cap q)$-Thomsen-Bedingung durch die (lediglich die Assoziativität der Addition besagende) $(P, Q, p \cap q)$-Reidemeister-Bedingung ersetzt werden kann. Da diese wegen der (P, p)-Transitivität gleichwertig ist mit der $(p \cap q, PQ)$-Transitivität (*PE*, S. 100, Satz 36), hätte man damit eine Kennzeichnung der desarguesschen Ebenen durch Transitivitätsaussagen, wodurch sich die Klasse II_3 in der Barlottischen Klasseneinteilung[7]) als leer ergeben würde. Hierfür gelang mir

[7]) Siehe Fußnote 4.

aber weder Beweis noch Widerlegung. Zum Schluß möchte ich noch darauf hinweisen, daß die im Beweis von Satz 2 erfolgte Übersetzung der (Q, q)-Transitivität in die Assoziativität der Verknüpfung $\circ$ nur die $(P, \mathfrak{p})$- und die $(\mathfrak{p} \cap q, PQ)$-Transitivität benutzt; denn es wurde (wie leicht nachzuprüfen ist) zwar die Assoziativität aber nicht die Kommutativität der Addition benutzt. Auch im Beweis von Satz 3 wurde die Kommutativität der Addition möglichst wenig gebraucht; daher zeigt der Beweis zugleich, daß unter den Bedingungen (a), (b), (c) bei Verzicht auf die Kommutativitätsforderung in (a) jedenfalls eine cartesische Gruppe mit assoziativer Multiplikation vorliegt.

Tübingen, Mathematisches Institut der Universität

(Eingegangen am 17. Oktober 1958)

ZELLER, K.
Math. Zeitschr. 71, 109–112 (1959)

Saturation bei äquivalenten Summierungsverfahren

Von
KARL ZELLER

§ 1. Einleitung

Bei der Saturation geht es darum, für einen Operationenfilter die höchstmögliche Konvergenzgeschwindigkeit zu bestimmen sowie diejenigen Elemente anzugeben, bei denen diese erreicht wird. Zum Beispiel erfüllen die C_1-Mittel σ_n der Fourier-Reihe einer Funktion f die Beziehung $\sup_x |\sigma_n(x) - f(x)| = O(1/n)$ für alle f einer leicht zu charakterisierenden „Saturationsklasse", während die entsprechende o-Abschätzung nur im Falle $f = 0$ möglich ist.

Der Begriff der Saturation wurde von FAVARD geprägt. ALEXITS, BUTZER und ZAMANSKY lösten eine Reihe von Saturationsproblemen (s. die Literaturangaben in [1]). BUTZER hat in der Arbeit [1] insbesondere darauf hingewiesen, daß sich äquivalente Summierungsverfahren bezüglich Saturation ganz verschieden verhalten können. Diese Erscheinung behandeln wir hier vom allgemeinen Standpunkt aus: Die Saturationsgeschwindigkeit bei Matrixverfahren hängt von der Konvergenzgeschwindigkeit der Spalten ab (Satz 1), letztere ist aber unabhängig von der Größe des Wirkfeldes (Satz 2).

§ 2. Definitionen und Bezeichnungen

Wir schließen uns bei Summierungsverfahren an die Bezeichnungsweise in [4] an. Alle bei uns vorkommenden Verfahren sind Matrixverfahren in Reihe-Folge-Form; eine Matrix A ordnet einer Reihe $\sum u_m$ folgendermaßen einen Grenzwert zu:

$$A\text{-}\sum u_m = \lim_{n\to\infty} \sum_{m=0}^{\infty} a_{nm} u_m,$$

wobei die Existenz der rechts stehenden Reihen für alle $n = 0, 1, \ldots$ vorausgesetzt wird. Folgende Eigenschaften von Matrizen spielen eine Rolle:

$$(\mathrm{Sp}_1) \qquad \lim_{n\to\infty} a_{nm} = 1 \qquad (m = 0, 1, \ldots),$$

$$(\mathrm{Sp}_{=1}) \qquad a_{nm} = 1 \qquad (n > n(m);\ m = 0, 1, \ldots),$$

$$(\mathrm{Zs}_{=1}) \qquad \sum_{m=0}^{\infty} a_{nm} = 1 \qquad (n = 0, 1, \ldots),$$

$$(\mathrm{Dr}) \qquad a_{nn} \neq 0, \quad a_{nm} = 0 \qquad (m > n;\ n = 0, 1, \ldots).$$

An einigen Stellen modifizieren wir $(\mathrm{Sp}_{=1})$ durch einen Zusatz dahingehend, daß die Bedingung nicht für alle, sondern nur für bestimmte m gefordert

wird. Weiter setzen wir

$$\|A\| = \sup_{n=0,1,\dots} \sum_{m=0}^{\infty} |a_{nm}|.$$

Ein BK-Raum $\mathfrak{E}$ ist ein B-Raum, der aus Zahlenfolgen $\mathfrak{x} = \{x_k\}$ besteht, wobei die Abbildungen $\mathfrak{x} \to x_k$ stetige Linearformen sind. Ein B-Raum $\mathfrak{E}$ ist genau dann als BK-Raum aufzufassen, wenn es in $\mathfrak{E}$ eine totale Folge stetiger Linearformen gibt. Lassen wir hier allgemeiner zu, daß die x_k in einem B-Raum $\mathfrak{B}$ liegen, so erhalten wir in analoger Weise den Begriff des BB-Raumes. Die Menge der Fourier-Reihen $\sum A_m(t)$ aller stetigen Funktionen können wir z. B. als BB-Raum $\mathfrak{E}$ auffassen; ein $\mathfrak{x} \in \mathfrak{E}$ hat die Funktionen A_m als Koordinaten.

§ 3. Allgemeine Saturation

Wir betrachten eine (Sp_1)-Matrix A und einen BB-Raum $\mathfrak{E}$, der — zwecks einfacherer Formulierung — mit einem $\mathfrak{x} \in \mathfrak{E}$ auch alle Elemente

$$\{\lambda_0 x_0, \dots, \lambda_m x_m, 0, \dots\}$$

enthalte (mit beliebigen Skalaren λ_i). Für $\mathfrak{x} \in \mathfrak{E}$ bilden wir

$$\mathfrak{y}^{(n)} = \{a_{nm} x_m\}_{m=0,1,\dots}.$$

Unter gewissen Bedingungen an A (etwa wenn A zeilenfinit ist) liegen die $\mathfrak{y}^{(n)}$ in $\mathfrak{E}$. Wegen (Sp_1) gilt $\mathfrak{y}^{(n)} \to \mathfrak{x}$ jedenfalls für gewisse $\mathfrak{x}$. Wir fragen: Was ist die bestmögliche Konvergenzgeschwindigkeit? (In etwas speziellerer Form wurde diese Frage inzwischen von FAVARD [2] behandelt.)

Als Beispiel nehmen wir für $\mathfrak{E}$ den obenerwähnten BB-Raum von Fourier-Reihen stetiger Funktionen, für A die C_1-Matrix (Reihe-Folge-Form); die $\mathfrak{y}^{(n)}$ stellen dann die Fejér-Mittel dar.

Da in einem BB-Raum die Koordinaten linear und stetig vom Element abhängen, gilt

$$\|x_m\| \cdot |1 - a_{nm}| = \|y_m^{(n)} - x_m\| = O(\|\mathfrak{y}^{(n)} - \mathfrak{x}\|) \quad (n \to \infty;\ m = 0, 1, \dots).$$

Eine unmittelbare Folgerung hieraus formulieren wir der Übersichtlichkeit halber als

Satz 1. *Gilt $x_m \neq 0$ (m fest), so ist*

$$|a_{nm} - 1| = O(\|\mathfrak{y}^{(n)} - \mathfrak{x}\|) \quad (n \to \infty),$$

d. h. die Konvergenzgeschwindigkeit von $\mathfrak{y}^{(n)} \to \mathfrak{y}$ ist nicht größer als die der m-ten Spalte von A.

Man sieht, daß es nicht immer eine höchste Konvergenzgeschwindigkeit gibt (z. B. wenn die Spalten mit wachsendem m immer rascher dem Wert 1 zustreben), und daß in gewissen Fällen die Saturationsklasse trivial ist (wenn eine Spalte rascher als alle anderen konvergiert). Am interessantesten ist der Fall gleicher Konvergenzgeschwindigkeit in allen Spalten, etwa

$$0 < \varrho_n = O(a_{nm} - 1) \quad \text{und} \quad (a_{nm} - 1) = O(\varrho_n) \quad (n \to \infty;\ m = 0, 1, \dots).$$

Die Saturationsklasse besteht nun aus den x mit

$$\left\| \left\{ \frac{a_{nm}-1}{\varrho_n} x_m \right\}_{m=0,1,\ldots} \right\| = O(1) \qquad (n \to \infty).$$

Von dieser Charakterisierung mittels Faktorfolgen ist natürlich noch ein weiter Weg zur funktionentheoretischen Bestimmung (im Fall von Fourier-Reihen). Unsere Betrachtungen lassen sich auch auf F-Räume ausdehnen, jedoch wird dann die Definition der Konvergenzgeschwindigkeit umständlicher.

§ 4. Summierungsverfahren

Wir betrachten nun Summierungsverfahren von dem in § 2 beschriebenen Typ (für die verwendeten Begriffe s. [4]). Die Bedingung (Sp_1) ist natürlich notwendig für die Permanenz eines solchen Verfahrens. Die Konvergenzgeschwindigkeit in den Spalten ist dagegen unabhängig von der Größe des Wirkfeldes. Durch Wiederholung von Zeilen drücken wir — abgesehen vom Fall (Sp_{-1}) — diese Geschwindigkeit beliebig herab, ohne das Wirkfeld zu verändern. Nach Satz 2 ist andererseits auch die höchstmögliche Geschwindigkeit — nämlich (Sp_{-1}) — erreichbar [wobei natürlich meistens die obengenannten $n(m)$ sehr rasch wachsen]. Die Beweismethoden zeigen, daß man die Geschwindigkeit auch in feinerer Weise variieren kann.

Satz 2. *Ist A eine (Sp_1)-Matrix, so gibt es zu dem Verfahren A ein äquivalentes und verträgliches Verfahren B, dessen Matrix (Sp_{-1}) erfüllt.*

Bevor wir zu dem vollen Beweis des Satzes schreiten, erläutern wir zwei Spezialfälle. Gibt es in A eine Zeile mit lauter Nichtnullen, so setzt man in jeder Spalte von A alle „genügend späten" Elemente gleich eins und erhält so das gewünschte B, weil die genannte Zeile allen A- oder B-summierbaren Reihen eine bestimmte Größenordnungsbedingung aufzwingt. Dasselbe Vorgehen führt auch bei einer (Dr)-Matrix A zum Ziel; man stützt sich dabei auf den unten benützten Mercersatz.

Dem Beweis von Satz 2 schicken wir einen Hilfssatz voraus.

Hilfssatz. *Sei A eine (Sp_1)-Matrix, m eine natürliche Zahl und $\varepsilon > 0$. Dann gibt es eine spaltenfinite* (Dr)*-Matrix C mit ($\mathrm{Zs}_{=1}$) und $\|C - I\| < \varepsilon$, bei der CA die Bedingung (Sp_{-1}) jedenfalls für das betreffende m erfüllt.*

Beweis. C wird so definiert, daß in jeder Zeile höchstens zwei Nichtnullen auftreten:

$$c_{k,m_k} = \varepsilon_k \qquad (\text{wo } 0 \leqq m_k < k;\ k = 0, 1, \ldots)$$

$$c_{kk} = 1 - \varepsilon_k \qquad (k = 0, 1, \ldots).$$

Dabei gelte

$$\varepsilon_k = 0 \qquad (k \leqq k^*),$$

$$\varepsilon_k a_{m_k, m} + (1 - \varepsilon_k)\, a_{km} = 1 \qquad (k > k^*).$$

Bei geeigneten $m_k \to \infty$ läßt sich dies mit $|\varepsilon_k| < \varepsilon/2$ erfüllen.

Wegen (Zs_{-1}) und der Spaltenfinitheit zerstört der Faktor C keine in A etwa schon in einer Spalte erfüllte Bedingung (Sp_{-1}). In der Form eines Produktes

$$C^{(m)} = C_m \dots C_0$$

erhalten wir daher eine spaltenfinite (Dr)-Matrix $C^{(m)}$ mit (Zs_{-1}) und $\|C^{(m)} - I\| < \varepsilon$, bei der $C^{(m)}A$ die Bedingung (Sp_{-1}) sogar in allen Spalten $0, 1, \dots, m$ erfüllt. Und aus den Zeilen dieser $C^{(m)}$ stellen wir weiter ein C zusammen, bei dem darüber hinaus CA der Bedingung (Sp_{-1}) sogar für alle m genügt. (Für diese Methode vgl. [3].)

Da nun C eine (Dr)-Matrix mit „überwiegender Hauptdiagonale" ist, folgt aus dem bekannten Mercer-Satz III auf S. 78 in [4], daß die Matrix $B = CA$ ein mit A äquivalentes und verträgliches Verfahren liefert, w.z.z.w.

Literatur

[1] BUTZER, P. L.: Zur Frage der Saturationsklassen singulärer Integraloperatoren. Math. Z. **70**, 93–112 (1958). — [2] FAVARD, J.: Sur la saturation des procédés de sommation. J. de Math. **36**, 359–372 (1957). — [3] WILANSKY, A., u. K. ZELLER: Summation of bounded divergent sequences, topological methods. Trans. Amer. Math. Soc. **78**, 501–509 (1955). — [4] ZELLER, K.: Theorie der Limitierungsverfahren. Heidelberg 1958.

Mathematisches Institut der Universität Tübingen

(Eingegangen am 21. Oktober 1958)

WAGNER, A.
Math. Zeitschr. 71, 113–123 (1959)

On perspectivities of finite projective planes

By
A. WAGNER

§ 1. Introduction

The main purpose of this note is to generalize two recent results in the theory of projective planes. These may be summarized in one theorem:

THEOREM A. *Let $\mathfrak{P}$ be a finite projective plane and Π a collineation group of $\mathfrak{P}$. If for every point C and line l of $\mathfrak{P}$, with $C \in l$ ($C \notin l$) there exists a non-trivial (C, l)-elation (homology) belonging to Π, then $\mathfrak{P}$ is desarguesian and Π contains the little projective group of $\mathfrak{P}$.*

The case where the collineation group contains elations is due to GLEASON [7], while the case where the collineation group contains homologies is due to ANDRÉ [1].

Essentially, Theorem A postulates the existence of elations (homologies) for all possible combinations of centres and axes. Our generalization consists in postulating instead that every point is the centre of a non-trivial elation (homology) and every line is the axis of a non-trivial elation (homology).

In the case of elations, we show in § 3 that the same conclusions as in Theorem A still follow from our weaker postulates.

The case where the collineation group contains homologies is treated in § 4. However, a more complicated situation arises here. We classify the resulting planes according to their collineation groups. These planes are necessarily translation planes or duals of translation planes.

That these planes need not be desarguesian we demonstrate by means of examples in § 5.

Some of the non-desarguesian planes used in § 5 have the additional property of containing ovals. Though this property is irrelevant to the rest of the note, we take the opportunity of demonstrating it in § 6.

Our previous results enable us to deduce a very simple condition for a finite plane to be desarguesian. We show in § 7 that if a finite plane has a collineation group transitive on the points of the plane and, in addition, possesses a non-trivial perspectivity, it is desarguesian.

The results of this note will be used in a subsequent paper [12], to which this note serves in some measure as an introduction. In both these papers we postulate the existence of collineation groups possessing certain properties. We are interested not only in deducing the configurational consequences of the collineation groups, but also in showing the nature of the postulated collineation groups. It is for this reason that we have tried to make clear

throughout within which collineation group we are working. We use group theoretical, particularly finite permutation group theoretical, methods to generate further collineations, particularly perspectivities, from the initially postulated collineations.

§ 2. Definitions and preliminary results

For definitions of the terms *projective plane*, *affine plane* and *collineation* see for example PICKERT [*14*]. It is well known that an affine plane may be embedded in a projective plane by adjoining a further line, often denoted by l_∞, and the points thereon. However, instead of regarding affine planes as separate entities from projective planes, we shall use the term *affine plane* to mean a projective plane in which one line is distinguished, or regarded as *special*. This line we shall denote by l_∞. The points of the plane not lying on l_∞ will be called *affine points*. The lines of the plane, except l_∞, will be called *affine lines*. To fit in with this convention, we restrict for affine planes the term collineation to those collineations which leave l_∞ fixed.

If a projective plane contains a finite number of points, the plane will be said to be *finite*. If a line of the plane contains $n+1$ points, every line of the plane contains $n+1$ points and there are $n+1$ lines through every point. The number n will be said to be the *order* of the plane.

Defining multiplication for collineations as the successive carrying out of collineations, collineations form a group. We shall read the product of collineations from left to right. Consequently, the image of a point P under a collineation α will be written as $P\alpha$.

A collineation of order 2 will be said to be *involutory*.

If a collineation σ of a projective plane leaves every point of a line l fixed, σ will be called a *perspectivity*. If σ is not the identity, there exists a unique point C such that every line through C is left fixed by σ. The line l will be called the *axis* of σ and the point C will be called the *centre* of σ. σ will be said to be a (C, l)-*perspectivity*. If $C \in l$, σ will be called an *elation*. If $C \notin l$, σ will be called a *homology*. To avoid tedious repetition, we shall henceforth, throughout this note, unless the contrary is indicated, use the term elation (homology) to signify a non-trivial elation (homology).

We draw attention to a few elementary properties of perspectivities (their duals are, of course, also true). The only points left fixed by a perspectivity are the centre and the points on the axis. The centre of an elation (homology) is the only point containing more than one (two) fixed lines. The perspectivities with given axis form a group. If σ is a (C, l)-perspectivity and ϱ is any collineation, then $\varrho^{-1}\sigma\varrho$ is a $(C\varrho, l\varrho)$-perspectivity.

The group generated by all the elations of a plane $\mathfrak{P}$ will be called the *little projective group* of $\mathfrak{P}$. The group generated by all the perspectivities of a plane $\mathfrak{P}$ will be called the *full projective group* of $\mathfrak{P}$.

An elation of an affine plane $\mathfrak{P}^*$ with axis l_∞ will be called a *translation*. We note that the product of two translations is a translation. The group

generated by all the translations of an affine plane will be called the *translation group* of the plane. If the translation group of $\mathfrak{P}^*$ is transitive on the affine points of $\mathfrak{P}^*$, then $\mathfrak{P}^*$ will be called a *translation plane*.

A projective plane $\mathfrak{P}$ will be said to be (C, l)-*transitive* (with respect to the collineation group Π) if for the fixed point C and fixed line l and any pair of points Q and Q', where $Q, Q' \notin l$, $QQ' \neq C$, $CQ = CQ'$, there exists a (C, l)-perspectivity ($\in \Pi$), with $Q\sigma = Q'$.

For a definition of the terms *alternative plane* and *desarguesian plane* see for example PICKERT [*14*] (alternative planes are there referred to as Moufang planes).

The following are some of the basic properties of projective planes which are subsequently used implicitly.

A plane is alternative if, and only if, it may be coordinatised by elements of an alternative field. A plane is also alternative if and only if the plane is (C, l)-transitive for all points C and lines l, with $C \in l$.

A plane is desarguesian if, and only if, it may be coordinatised by elements of a field (not necessarily commutative). A plane is also desarguesian if, and only if, the plane is (C, l)-transitive for all points C and lines l.

Since every finite alternative field is a field, every finite alternative plane is desarguesian.

In this note we shall have repeated occasion to use the following theorem due to ANDRÉ [*1*]:

THEOREM B. *Let $\mathfrak{P}^*$ be a finite affine plane and let σ_1 and σ_2 be two homologies of $\mathfrak{P}^*$ with centres C_1 and C_2 respectively and common axis l_∞. Then the group generated by σ_1 and σ_2 contains an elation mapping C_1 onto C_2.*

COROLLARY (i). *Let m be an affine line of $\mathfrak{P}^*$ and Π a collineation group of $\mathfrak{P}^*$ such that for every affine point $C \in m$ there exists a (C, l_∞)-homology. Then $\mathfrak{P}^*$ is $(m \cap l_\infty, l_\infty)$-transitive with respect to Π.*

COROLLARY (ii). *Let Π be a collineation group of $\mathfrak{P}^*$ such that for every affine point C there exists a (C, l_∞)-homology belonging to Π. Then $\mathfrak{P}^*$ is a translation plane and Π contains the translation group of $\mathfrak{P}^*$ as a subgroup.*

It is easily seen that the case of Theorem A dealing with homologies is essentially a further corollary of Theorem B.

In § 4 and § 6 we shall also require the following concepts.

We define a *correlation* of a projective plane $\mathfrak{P}$ as a one-to-one mapping which maps the points of $\mathfrak{P}$ onto the lines of $\mathfrak{P}$ and the lines of $\mathfrak{P}$ onto the points of $\mathfrak{P}$ such that a point is incident with a line if, and only if, the corresponding image line is incident with the corresponding image point. Clearly the successive carrying out of two correlations is a collineation. A correlation ϑ will be said to be a *polarity* if ϑ^2 is the identity collineation. A point Q will be said to be an *absolute point* of the polarity ϑ if $Q \in Q\vartheta$.

If $\mathfrak{P}$ is a finite projective plane of odd order n, then a set of $n+1$ points such that no three are collinear will be called an *oval*.

§ 3. Collineation groups containing elations

THEOREM 1. *Let $\mathfrak{P}$ be a finite projective plane and let Π be a collineation group of $\mathfrak{P}$ with the following properties:*

(i) For every point $C \in \mathfrak{P}$ there exists an element $\lambda \in \Pi$ such that λ is an elation with centre C.

(ii) For every line $l \in \mathfrak{P}$ there exists an element $\varrho \in \Pi$ such that ϱ is an elation with axis l. Then $\mathfrak{P}$ is desarguesian and Π contains the little projective group of $\mathfrak{P}$ as a subgroup.

Let us assume that there exists in $\mathfrak{P}$ a point C and a line l, with $C \in l$, such that Π does not contain a (C, l)-elation. We denote by $D_1, \ldots, D_s$, with $D_i \neq D_j$ if $i \neq j$, the centres of all the elations in Π which have axis l. Similarly, we denote by $m_1, \ldots, m_t$, with $m_i \neq m_j$ if $i \neq j$, the axes of all the elations which have centre C.

From the hypothesis it follows that $D_i \neq C$ $(i = 1, \ldots, s)$ and $m_j \neq l$ $(j = 1, \ldots, t)$. Now for each D_i $(i = 1, \ldots, s)$ denote by λ_i a (D_i, l)-elation in Π. Similarly, for each m_j $(j = 1, \ldots, t)$ denote by ϱ_j a (C, m_j)-elation in Π. Then λ_i cannot leave any line through C, except of course l, fixed. Also $\lambda_i^{-1} \varrho_1 \lambda_i$ is a $(C, m_1 \lambda_i)$-elation. Consequently $m_1 \lambda_i = m_u$, where u belongs to the set $2, \ldots, t$.

We shall now show that if $D_i \neq D_j$, then $m_1 \lambda_i \neq m_1 \lambda_j$. If $m_1 \lambda_i = m_1 \lambda_j$, then $m_1 \lambda_i \lambda_j^{-1} = m_1$, which would imply that $\lambda_i \lambda_j^{-1}$ is a (C, l)-elation, which contradicts our hypothesis. Consequently to the s points D_i $(i = 1, \ldots, s)$ there correspond s distinct lines m_u where u belongs to the set $2, \ldots, t$. Hence

$$s < t.$$

We now carry out an argument dual to the above and obtain

$$t > s.$$

This contradiction shows that for every point C and line l belonging to $\mathfrak{P}$, with $C \in l$, there exists a (C, l)-elation. From Theorem A the result follows.

§ 4. Collineation groups containing homologies

THEOREM 2. *Let $\mathfrak{P}$ be a finite projective plane and Π a collineation group of $\mathfrak{P}$ with the following properties:*

(i) For every point $C \in \mathfrak{P}$ there exists an element $\lambda \in \Pi$ such that λ is a homology with centre C.

(ii) For every line $l \in \mathfrak{P}$ there exists an element $\varrho \in \Pi$ such that ϱ is a homology with axis l.

Then the following cases[1]) exhaust the possibilities for $\mathfrak{P}$ and Π:

Case (a) Π fixes a line of $\mathfrak{P}$. Regarding the line fixed by Π as the special line of $\mathfrak{P}$, then $\mathfrak{P}$ is a translation plane and Π contains the translation group of $\mathfrak{P}$ as a proper subgroup.

[1]) Cases (a) and (b) are not mutually exclusive, see § 5.

Case (b) Π fixes a point of $\mathfrak{P}$. The conclusions are the dual of the conclusions of case (a).

Case (c) Π fixes no point and no line of $\mathfrak{P}$. Then $\mathfrak{P}$ is a desarguesian plane and Π contains the little projective group of $\mathfrak{P}$ as a subgroup.

To avoid tedious repetition, all collineations referred to in the proof of this Theorem will be understood to be elements of Π. Further, we shall call the line l an *axis of the point* C if there exists a (C, l)-homology. Similarly, we shall call C a *centre of the line* l if there exists a (C, l)-homology.

The following argument will be used repeatedly in the proof:

If the line l is an axis of the point C and σ is any collineation of the plane, then $l\sigma$ is an axis of $C\sigma$.

This is an immediate consequence of our earlier remarks concerning the conjugates of perspectivities. The dual of the above argument is clearly also true.

We shall now prove Theorem 2. We first observe that properties (i) and (ii) are mutually dual.

Case (a). Π fixes a line of $\mathfrak{P}$.

We denote the line fixed by Π by l_∞ and regard $\mathfrak{P}$ as an affine plane with special line l_∞. The axis of every affine point is obviously l_∞. Consequently, for every affine point C there exists a homology with centre C and axis l_∞. The conclusions for case (a) now follow immediately from the corollary (ii) of Theorem B.

Case (b). Π fixes a point of $\mathfrak{P}$.

An argument dual to the argument for case (a) is used.

Case (c). Π fixes no point and no line of $\mathfrak{P}$.

We shall first prove three statements.

STATEMENT I. *Some point of $\mathfrak{P}$ has more than one axis or some line of $\mathfrak{P}$ has more than one centre.*

Let us assume the above to be false. This implies that every point of $\mathfrak{P}$ has one and only one axis and every line of $\mathfrak{P}$ has one and only one centre.

We now define a mapping ϑ of the points of $\mathfrak{P}$ onto the lines of $\mathfrak{P}$ and of the lines of $\mathfrak{P}$ onto the points of $\mathfrak{P}$ as follows:

For every point C of $\mathfrak{P}$, $C\vartheta$ = the axis of C.

For every line l of $\mathfrak{P}$, $l\vartheta$ = the centre of l.

We now assert that ϑ is a correlation. To prove this we must show that if C is any point and l any line of $\mathfrak{P}$, then $C \in l$ if, and only if, $l\vartheta \in C\vartheta$. Since ϑ is clearly an involutory mapping, it will be sufficient to show that if $C \in l$, then $l\vartheta \in C\vartheta$. Let λ be a $(C, C\vartheta)$-homology. Then $(l\vartheta)\lambda$ is the centre of $l\lambda$. Since $C \in l$, it follows that $l\lambda = l$, hence $(l\vartheta)\lambda = l\vartheta$. But the only points of $\mathfrak{P}$ left fixed by λ are C and the points of $C\vartheta$. Since $C \in l$ and $l\vartheta \notin l$, it follows that $l\vartheta \neq C$. Hence $l\vartheta \in C\vartheta$, which establishes that ϑ is a correlation.

We have shown that ϑ is a polarity. Our initial assumption implies that ϑ has no absolute points. However, it has been shown by BAER [4] that every polarity of a finite plane possesses absolute points. This contradiction proves Statement I.

STATEMENT II. *Let some line of $\mathfrak{P}$ possess more than one centre but no line of $\mathfrak{P}$ possess three non-collinear centres. Then some point of $\mathfrak{P}$ has three non-concurrent axes.*

Let l be a line of $\mathfrak{P}$ possessing more than one centre. Denote two of these centres by C_1 and C_2. Further, denote the line C_1C_2 by m and let $E = l \cap m$. Now let D be any point of l except E, and let λ be a homology with centre D. Since $l\lambda = l$, it follows that $C_1\lambda$ and $C_2\lambda$ are centres of l. Consequently $C_1\lambda$, $C_2\lambda \in m$. This means that m is left fixed by λ. However, the only lines of $\mathfrak{P}$ left fixed by λ are the lines through D and the axis of λ. $D \notin m$, hence m is the axis of D. Consequently, every point of l, except E, is a centre of m. From Theorem B corollary (i), it follows that $\mathfrak{P}$ is (E, m)-transitive.

l contains at least two centres of m. Consequently, with the rôles of l and m interchanged, an argument similar to the above shows that $\mathfrak{P}$ is (E, l)-transitive. Now the group of collineations generated by all elations with centre E and axis l or m leaves E fixed and is transitive on the lines of $\mathfrak{P}$ not passing through E (see PICKERT [*14*], p. 67, dual of Theorem 11). The axis of E does not pass through E, hence every line not passing through E is an axis of E. This establishes Statement II.

STATEMENT III. *If some line of $\mathfrak{P}$ possesses three non-collinear centres, then $\mathfrak{P}$ is desarguesian and Π contains the little projective group of $\mathfrak{P}$.*

Let l_1 be a line of $\mathfrak{P}$ possessing three non-collinear centres, denoted by C_1, C_2, C_3. Let C_1C_3 meet l_1 in D_1 and C_2C_3 meet l_1 in D_2. Obviously $D_1 \neq D_2$. From Theorem B, it follows that there exists a (D_1, l_1)- and a (D_2, l_1)-elation; denote these by ϱ_1 and ϱ_2 respectively.

Now let D be any point of l_1, with $D \neq D_1, D_2$. Let m be an axis of D. m cannot contain both D_1 and D_2. Let us suppose that $D_1 \notin m$. Since $D \in l_1$, $D\varrho_1 = D$. Consequently $m\varrho_1$ is an axis of D. Since $D_1 \notin m$, $m\varrho_1 \neq m$. We may now use the dual of Theorem B to show that there exists an elation with centre D which carries m into $m\varrho_1$. Since m and $m\varrho_1$ meet on l_1, it follows that l_1 is the axis of this elation. Hence for every point D, with $D \in l_1$, there exists a (D, l_1)-elation.

Every transitive class of lines under Π contains at least three non-concurrent lines. This is easily seen as follows. Let h_1 and h_2 be two distinct lines in the same transitive class under Π, and let $h_1 \cap h_2 = P$. Further, let π be an element of Π such that $P\pi = P$. Then either $h_1, h_2, h_1\pi$ or $h_1, h_2, h_2\pi$ are three non-concurrent lines.

Now let l_2 and l_3 be any two lines of $\mathfrak{P}$ in the same transitive class of lines under Π as l_1, such that l_1, l_2, l_3 are non-concurrent. Then the following is true for $i = 1, 2, 3$: if a point $D \in l_i$, there exists a (D, l_i)-elation. We shall refer to this property as Property 1.

We shall now show that Property 1 implies that every line of $\mathfrak{P}$ is the axis of an elation.

Let m be any line of $\mathfrak{P}$ and let σ be a homology with m as axis. Further, let C be the centre of σ. C cannot lie on l_1, l_2, l_3. We assume that $C \notin l_1$.

Let $D = m \cap l_1$ (if $l_1 = m$, then choose any point $D \in l_1$). Also let ϱ be a (D, l_1)-elation. Clearly $m\varrho = m$ and $C\varrho \neq C$. Hence there exists a $(C\varrho, m)$-homology, and from Theorem B it follows that there exists a (D, m)-elation.

Now let $E_1 = l_2 \cap l_3$, $E_2 = l_3 \cap l_1$, $E_3 = l_1 \cap l_2$. We shall now examine the case where the line m passes through E_i ($i = 1$, 2 or 3). Then an argument similar to the above shows that there exists a (E_i, m)-elation. Thus we have shown that the dual of Property 1 also holds, and the dual of the preceding argument shows that every point of $\mathfrak{P}$ is the centre of some elation. Appeal to Theorem 1 completes the proof of Statement III.

We remark that by dual arguments the duals of Statements II and III may be established.

We are now able to complete the proof for case (c). Statement I shows that either some line of $\mathfrak{P}$ has more than one centre or some point of $\mathfrak{P}$ has more than one axis. Let us first assume that the former alternative is the case. Then if some line of $\mathfrak{P}$ has three non-collinear centres, Statement III completes the proof. If, however, no line of $\mathfrak{P}$ has three non-collinear centres, then Statement II shows that some point of $\mathfrak{P}$ has three non-concurrent axes. Then the dual of Statement III completes the proof. Finally, if the second alternative of Statement I is the case, then the dual of the preceding argument completes the proof.

§ 5. Examples of non-desarguesian planes satisfying Theorem 2

The question of the existence of non-desarguesian planes satisfying the conditions of Theorem 2 naturally arises. That such planes do exist we shall now demonstrate by means of examples.

Let $\mathscr{A}$ denote a Veblen-Wedderburn algebra (for a definition see, for example, PICKERT [*14*] pp. 90–92, where such an algebra is referred to as a quasi-field). We assume the left distributive law to hold, that is

$$(5.1) \qquad (a + b)\, c = a\, c + b\, c \qquad \text{for all } a, b, c \in \mathscr{A}$$

from this it follows at once that

$$(5.2) \qquad (-a)\, b = -a\, b.$$

From the algebra $\mathscr{A}$ we may now construct a projective plane, denoted by $\mathfrak{P}(\mathscr{A})$ as follows (see for example VEBLEN and WEDDERBURN [*17*]):

(5.3) Points are the following triplets (x_1, x_2, x_3):
the "points at infinity" $(0, 1, 0)$ and $(1, a, 0)$ for all $a \in \mathscr{A}$,
the "affine points" $(a, b, 1)$ for all $a, b \in \mathscr{A}$.

(5.4) Lines are the following triplets $[u_1, u_2, u_3]$:
the "line at infinity" $[0, 0, 1]$,
the "affine lines" $[1, 0, a]$ and $[a, 1, b]$ for all $a, b \in \mathscr{A}$.

(5.5) The point (x_1, x_2, x_3) is incident with the line $[u_1, u_2, u_3]$ if, and only if,

$$x_1 u_1 + x_2 u_2 + x_3 u_3 = 0.$$

$\mathfrak{P}(\mathscr{A})$ is a translation plane with special line $[0, 0, 1]$ (see for example PICKERT [14] p. 108).

We shall henceforth assume the characteristic of $\mathscr{A}$ to be distinct from 2. (The characteristic of $\mathscr{A}$ is defined to be the order of 1 under addition. From (5.1) it follows that the order of all elements of $\mathscr{A}$ under addition is the same.)

The following mapping, α, is easily verified to be an involutory homology with centre $(0, 0, 1)$ and axis $[0, 0, 1]$:

$$\begin{aligned}
&(0, 1, 0) \to (0, 1, 0)\\
&(1, a, 0) \to (1, a, 0)\\
&(a, b, 1) \to (-a, -b, 1)\\
&[0, 0, 1] \to [0, 0, 1]\\
&[1, 0, a] \to [1, 0, -a]\\
&[a, 1, b] \to [a, 1, -b].
\end{aligned}$$

Since the collineation group of $\mathfrak{P}(\mathscr{A})$ is transitive on affine points, every affine point of $\mathfrak{P}(\mathscr{A})$ is the centre of an involutory homology.

We now assume that $\mathscr{A}$ has the further property:

$$a(-b) = -ab \quad \text{for all} \quad a, b \in \mathscr{A}. \tag{5.6}$$

Then the following mapping, β, is an involutory homology with centre $(0, 1, 0)$ and axis $[0, 1, 0]$:

$$\begin{aligned}
&(0, 1, 0) \to (0, 1, 0)\\
&(1, a, 0) \to (1, -a, 0)\\
&(a, b, 1) \to (a, -b, 1)\\
&[0, 0, 1] \to [0, 0, 1]\\
&[1, 0, a] \to [1, 0, a]\\
&[a, 1, b] \to [-a, 1, -b].
\end{aligned}$$

Further, let $\gamma = \alpha\beta$. γ is an involutory homology with centre $(1, 0, 0)$ and axis $[1, 0, 0]$.

It has been shown by ANDRÉ [2] that the near field with 9 elements, which we shall denote by $\mathscr{A}_9$, satisfies (5.6) and that the collineation group of $\mathfrak{P}(\mathscr{A}_9)$ is transitive on affine lines. Hence by considering conjugates of β we see that every affine line of $\mathfrak{P}(\mathscr{A}_9)$ is the axis and every point on the line at infinity is the centre of an involutory homology. Taking Π to be the group of all collineations of $\mathfrak{P}(\mathscr{A}_9)$, it now follows that $\mathfrak{P}(\mathscr{A}_9)$ is a non-desarguesian plane satisfying the conditions of Theorem 2. Any plane constructed from a proper near field cannot be both a translation plane and the dual of a translation plane. Thus $\mathfrak{P}(\mathscr{A}_9)$ has the properties of case (a) of Theorem 2 but not those of case (b).

The dual of $\mathfrak{P}(\mathscr{A}_9)$ is obviously a plane having the properties of case (b) but not those of case (a).

It is quite possible for a plane to be an example of case (a) and case (b) simultaneously, as we shall now demonstrate.

Let $\mathscr{A}^*$ be a finite Veblen-Wedderburn algebra, not a field, in which both distributive laws hold. Examples of such algebras are given in DICKSON [6]. Since the right distributive law holds, (5.6) is satisfied. The plane $\mathfrak{P}(\mathscr{A}^*)$ is of course a translation plane; moreover it is the dual of a translation plane with special point (0, 1, 0) (see for example PICKERT [14] p. 101, Theorem 39). Denoting (0, 1, 0) by C, this implies that $\mathfrak{P}(\mathscr{A}^*)$ is (C, l)-transitive, where l is any line through C. The collineation group of $\mathfrak{P}(\mathscr{A}^*)$ leaves the line at infinity and C fixed. It is clearly transitive on the affine lines not through C, hence also on the points on the line at infinity distinct from C. By considering conjugates of β we see that every affine line not through C is the axis of an involutory homology, the centre of these homologies being C. By considering conjugates of γ we see that every point on the line at infinity distinct from C is the centre and every affine line through C is the axis of an involutory homology. Taking Π to be the group of all collineations of $\mathfrak{P}(\mathscr{A}^*)$, $\mathfrak{P}(\mathscr{A}^*)$ satisfies the conditions of Theorem 2 and has the properties of case (a) and case (b).

§ 6. Examples of ovals in non-desarguesian planes

Ovals have received considerable study (see for example BAER [4], OSTROM [10], QVIST [15], SEGRE [16]). The only previously given examples known to the author of planes possessing ovals are the desarguesian planes and the cyclic planes. No finite non-desarguesian cyclic planes are known. Certain of the planes $\mathfrak{P}(\mathscr{A}^*)$, discussed in § 5 provide ready, and obvious, examples of non-desarguesian planes possessing ovals.

Let $\mathscr{A}^{**}$ be a commutative Veblen-Wedderburn algebra of odd characteristic. The two-sided distributive Veblen-Wedderburn algebras due to Dickson referred to in § 5 are all of this type. We note that every element of the subfield generated by 0 and 1, the additive and multiplicative units of $\mathscr{A}^{**}$, associate with all elements of $\mathscr{A}^{**}$.

$\mathfrak{P}(\mathscr{A}^{**})$ has a polarity given by

$$(0, 1, 0) \leftrightarrow [0, 0, 1]$$
$$(1, a, 0) \leftrightarrow [1, 0, a]$$
$$(a, b, 1) \leftrightarrow [a, 1, b].$$

The absolute points of this polarity are readily seen to be (0, 1, 0) and $(a, -\frac{1}{2}a^2, 1)$ for all $a \in \mathscr{A}^{**}$. That these points form an oval may easily be verified directly. Alternatively, we may use a result of BAER [4], Theorem 5, Corollary 1, which states that if the order of a plane is odd and the number of absolute points is equal to the number of points on each line, then the absolute points form an oval.

We remark that we have not succeeded in demonstrating the existence of ovals which are not the absolute points of some polarity.

§ 7. Transitive projective planes

We shall call a projective plane $\mathfrak{P}$ *transitive* if the collineation group of $\mathfrak{P}$ is transitive on the points of $\mathfrak{P}$. We shall restrict our attention to transitive finite projective planes.

We have defined a transitive projective plane by referring the effect of the collineation group to points. In the case of finite planes it is immaterial whether transitivity is referred to points or lines. This is an immediate consequence of a more general result due, independently, to DEMBOWSKI [5], HUGHES [9], and PARKER [13], which states that for any collineation group of a finite projective plane the number of transitive classes of points is equal to the number of transitive classes of lines.

It is not known whether every transitive finite plane is desarguesian. We shall prove a sufficient condition for a transitive finite plane to be desarguesian. It is quite obvious by considering, for example, the little projective group of a desarguesian plane, that this condition is also necessary.

THEOREM 3. *Let $\mathfrak{P}$ be a finite projective plane and let Π be a collineation group of $\mathfrak{P}$ such that Π is transitive on the points of $\mathfrak{P}$ and Π contains a nontrivial perspectivity. Then $\mathfrak{P}$ is desarguesian and Π contains the little projective group of $\mathfrak{P}$.*

Let the given perspectivity be denoted by σ. Let the centre and axis of σ be denoted by C and l respectively. We shall treat separately the case where σ is an elation and the case where σ is a homology.

Case (i) σ is an elation.

Let D be any point of $\mathfrak{P}$. Since Π is transitive on the points of $\mathfrak{P}$ there exists an element $\pi \in \Pi$ such that $C\pi = D$. Then D is the centre of the elation $\pi^{-1}\sigma\pi$. Since Π is also transitive on the lines of $\mathfrak{P}$, a similar argument may be used to show that every line of $\mathfrak{P}$ is the axis of a non-trivial elation. The result now follows immediately from Theorem 1.

Case (ii) σ is a homology.

An argument similar to the above may be used to show that every point of $\mathfrak{P}$ is the centre of a non-trivial homology and every line of $\mathfrak{P}$ is the axis of a non-trivial homology. From Theorem 2 it follows that Π contains an elation. The argument of case (i) above may now be used to complete the proof.

COROLLARY. *Let $\mathfrak{P}$ be a finite projective plane and Π a collineation group of $\mathfrak{P}$ transitive on the points of $\mathfrak{P}$. If the order of $\mathfrak{P}$ is not a square and the order of Π is even, then $\mathfrak{P}$ is desarguesian and Π contains the little projective group of $\mathfrak{P}$ as a subgroup.*

Since the order of Π is even, Π contains an involution. It was shown by BAER [3], that if the order of a finite plane is not a square, every involutory collineation is a perspectivity. We may now apply Theorem 3.

The order of a collineation group doubly transitive on the points of a finite plane is necessarily even. Thus as a special case of the above corollary we obtain the result originally due to OSTROM [11], and MARSHALL HALL and HUGHES [8], that every finite doubly transitive projective plane whose order is not a square is desarguesian.

References

[1] ANDRÉ, J.: Über Perspektivitäten in endlichen projektiven Ebenen. Arch. Math. **6**, 29–32 (1954). — [2] ANDRÉ, J.: Projektive Ebenen über Fastkörpern. Math. Z. **62** 137–160 (1955). — [3] BAER, R.: Projectivities with fixed points on every line of the plane. Bull. Amer. Math. Soc. **52**, 273–286 (1946). — [4] BAER, R.: Polarities in finite projective planes. Bull. Amer. Math. Soc. **52**, 77–93 (1946). — [5] DEMBOWSKI, H. P.: Verallgemeinerungen von Transitivitätsklassen endlicher projektiver Ebenen. Math. Z. **69**, 59–89 (1958). — [6] DICKSON, L. E.: On finite algebras. Nachr. Ges. Wiss. Göttingen, Math.-phys. Kl. **1905**, 358–394. — [7] GLEASON, A. M.: Finite Fano planes. Amer. J. Math. **78**, 797–807 (1956). — [8] HUGHES, D. R.: Generalized incidence matrices over group algebras. Illinois J. Math. **1**, 545–551 (1957). — [9] HUGHES, D. R.: Collineations and generalized incidence matrices. Trans. Amer. Math. Soc. **86**, 284–296 (1957). — [10] OSTROM, T. G.: Ovals, dualities, and DESARGUES's theorems. Canad. J. Math. **7**, 417–431 (1955). — [11] OSTROM, T. G.: Double transitivity in finite projective planes. Canad. J. Math. **8**, 563–567 (1956). — [12] OSTROM, T. G., and A. WAGNER: On projective and affine planes with transitive collineation groups. To appear in Math. Z. — [13] PARKER, E. T.: On collineations of symmetric designs. Proc. Amer. Math. Soc. **8**, 350–351 (1957). — [14] PICKERT, G.: Projektive Ebenen. Heidelberg 1955. — [15] QVIST, B.: Some remarks concerning curves of the second degree in a finite plane. Ann. Acad. Sci. Fennicae (A.I.) **134**, 12pp. (1952). — [16] SEGRE, B.: Ovals in a finite projective plane. Canad. J. Math. **7**, 414–416 (1955). — [17] VEBLEN, O., and J. H. M. WEDDERBURN: Non-desarguesian and non-pascalian geometries. Trans. Amer. Math. Soc. **8**, 379–388 (1907).

Dept. of Math., The Ohio State University, Columbus, Ohio (U.S.A.)
Imperial College, London

(Eingegangen am 26. August 1958)

HADWIGER, H.
Math. Zeitschr. 71, 124–140 (1959)

Normale Körper im euklidischen Raum und ihre topologischen und metrischen Eigenschaften*)

Von
H. HADWIGER

So leicht es auch ist, die aus den Gestalten der wirklich existierenden materiellen Dinge durch Abstraktion gewonnene Idee des normalen geometrischen Körpers rein intuitiv zu erfassen, so schwierig ist es seltsamerweise, diesen, unserer geometrischen Anschauung direkt verpflichteten, aber doch recht vagen Begriff auf zufriedenstellende Art mathematisch streng zu charakterisieren. In verschiedenen Sachgebieten stellt sich da und dort das Bedürfnis ein, sich auf eine Klasse ausreichend vernünftig gestalteter Körper des euklidischen Raumes beschränken zu können. Damit verbindet sich auch die Erwartung, daß sich bei dieser Beschränkung keine lediglich durch die Körpergestalt bedingten Schwierigkeiten einstellen, wie diese leider oft eine flüssige und in sich geschlossene Entwicklung innerhalb des betreffenden Sachbereichs beeinträchtigen oder verunmöglichen.

Wie können die lästigen pathologischen Entartungs- und Sonderfälle vermieden werden, wenn man sich nicht auf eine triviale Figurenklasse beschränken will, sondern bestrebt ist, jede einigermaßen vernünftige geometrische Form zuzulassen? Mit anderen Worten: Was ist ein normaler Körper? Unsere Frage hängt auch mit einem Hauptproblem der anschaulichen Topologie zusammen, wie es insbesondere von K. BORSUK [vgl. K. BORSUK, Über einige Probleme der anschaulichen Topologie, Jahresber. D. M. V. **60**, 101–114 (1958)] klar formuliert und erörtert worden ist. In unserem Fall handelt es sich um ein spezielleres und einfacheres Problem, indem es lediglich darum geht, eine Normalität für Punktmengen gewöhnlicher Räume, also für Körper im naiven Sinn, festzulegen. Das Dilemma, das sich bei Versuchen dieser Art einstellt, ist allgemein bekannt: Entweder ist die Körperklasse zu eng, um befriedigen zu können, was beispielsweise bei Beschränkung auf euklidische Polyeder jedenfalls zutrifft, oder es wird doch wieder ein zu weiter Spielraum für abnormale Erscheinungen offen gelassen, wie dies der Fall ist, wenn man die krummen Polyeder, also die topologischen Bilder euklidischer Polyeder, zuläßt.

Starke zusätzliche Bedingungen bezüglich der verwendeten Abbildung bringen unnatürliche und zum Teil wieder unerwünschte Einengungen, ohne daß das Problem befriedigend gelöst wird. Auch eine durch ein analytisches

*) Vortrag im Mathematischen Kolloquium der Universität Heidelberg am 27. Juni 1958.

Gesetz bedingte Gestalt kann Anomalien aufweisen, die man gerne aus dem Bereich des Zulässigen verbannt sähe. Differenzierbarkeitsvoraussetzungen erzwingen lediglich eine gewisse lokale gestaltliche Normalität, während die hier in Frage stehende Körpereigenschaft offenbar einen lokalen und globalen Charakter hat.

Allgemein darf gesagt werden, daß das sonst außerordentlich fruchtbare Abbildungsprinzip, dessen Einsatz in geometrischen Sachgebieten in vielen einschlägigen Fällen einen bedeutungsvollen Anschluß an die Analysis herstellt, im Rahmen unserer Problemstellung sich nicht besonders gut zu eignen scheint. Die unmittelbar in der geometrischen Anschauung verwurzelte Vorstellung vom normalen Körper findet die ihr adäquate Deutung eher innerhalb der direkten Mengengeometrie.

In der vorliegenden Abhandlung wird nun der Versuch unternommen, eine Klasse normaler Körper so zu kennzeichnen, daß hierbei sowohl metrische, auf der Euklidizität des Raumes beruhende Bedingungen, als auch solche topologischer Art, die sich auf den Zusammenhang beziehen, zur Geltung gebracht werden. Die Normalität soll also vom topologischen und vom metrischen Gesichtspunkt aus erfaßt werden. Das Zusammenwirken der beiden genannten, in gewissem Sinne entgegengesetzten Betrachtungsweisen bei der Lösung unserer Aufgabe ist im Hinblick auf die mannigfaltigen, zwischen Topologie und metrischer Geometrie bestehenden engen Beziehungen, wie diese besonders von H. Hopf wiederholt hervorgehoben und erörtert wurden [vgl. H. Hopf, Über Zusammenhänge zwischen Topologie und Metrik im Rahmen der elementaren Geometrie, Math. Phys. Semesterber. **3**, 16—29 (1953)], durchaus verständlich. Diese Wechselwirkungen ermöglichen es, die Normalität einem Gleichgewicht entsprechen zu lassen, das zwischen den auf die Körpergestaltung einwirkenden, entgegengesetzt gerichteten Kräften bestehen soll.

Bei Beschränkung auf die Klasse der so charakterisierten normalen Körper wird es möglich, die einfachsten Begriffe und Ergebnisse der elementaren kombinatorischen Topologie, soweit sich diese lediglich auf die als Punktmengen aufgefaßten Körper des gewöhnlichen Raumes beziehen, auf eine direkte und insbesondere von den üblichen, vornehmlich algebraischen Methoden der konventionellen Topologie unabhängige Art zu entwickeln. Im Mittelpunkt steht eine direkte Begründung der Charakteristik von Euler-Poincaré. Dieses Programm wird im I. Teil durchgeführt. Im anschließenden II. Teil wird gezeigt, daß für normale Körper die fundamentalen Maßzahlen, insbesondere die Quermaßintegrale Minkowskis in einheitlicher Weise als Integrale vom Croftonschen Typ definiert werden können, die sich über die Charakteristik von Schnittkörpern erstrecken. Die Normalität der Körper bewirkt hier, daß alle diese Ansätze als Riemannsche eigentliche Integrale existieren. Die vollständigen integralgeometrischen Formelsysteme lassen den engen Zusammenhang erkennen, der bei den normalen Körpern zwischen den durch die Charakteristik erfaßten topologischen Gegebenheiten einerseits und den metrischen Maßzahlen andererseits bestehen.

I. Teil

1. Raum. Es bezeichne E (genauer E_k, $k \geqq 0$) den k-dimensionalen euklidischen Raum und E' (genauer E_i, $0 \leqq i \leqq k$) bedeute einen i-dimensionalen euklidischen Teilraum von E.

2. Mehrfachrichtung. Unter einer Mehrfachrichtung (abgekürzt Mr) r im E verstehen wir ein System $r = \langle r_1, \ldots, r_k \rangle$ von k unabhängigen Richtungen, die durch k linear unabhängige Einheitsvektoren r_ν ($\nu = 1, \ldots, k$) gekennzeichnet sind. Ein i-dimensionaler Teilraum $E' = E_i$ heißt der Mr r assoziiert, wenn sich unter den k Richtungen r_ν ($\nu = 1, \ldots, k$) von r genau i passend ausgewählte Richtungen r'_μ ($\mu = 1, \ldots, i$) finden, die zu E' parallel sind. Diese bilden offenbar eine Mr $r' = \langle r'_1, \ldots, r'_i \rangle$ im $E' \cdot E_0$ und E_k sind jeder Mr r assoziiert.

3. Parzelle. Eine kompakte Punktmenge P von E heißt eine der Mr r im E assoziierte Parzelle $P = P(r)$, wenn der Durchschnitt $P \cap E'$ für jeden der Mr r assoziierten Teilraum E' zusammenhängend ist. Die leere Menge 0 gilt als kompakt und zusammenhängend; insbesondere ist die leere Menge 0 also eine jeder Mr r assoziierte Parzelle. Ein eigentlicher oder uneigentlicher konvexer Körper ist eine Parzelle, die jeder Mr r assoziiert ist; dadurch ist ein solcher übrigens auch charakterisiert.

4. Schnitteigenschaft der Parzellen. Ist $P = P(r)$ eine der Mr r assoziierte Parzelle und ist E' ein ebenfalls der Mr r assoziierter Teilraum, so ist der Durchschnitt $P' = P \cap E'$ wieder eine der Mr r assoziierte Parzelle $P' = P'(r)$. Dies folgt unmittelbar aus den Definitionen. Ist r' die beigeordnete Mr im E', d.h. die von den zu E' parallelen Richtungen von r im E gebildet wird, so ist P' als Punktmenge im E' betrachtet eine dort der Mr r' assoziierte Parzelle $P' = P'(r')$.

5. Parzellierung. Eine Punktmenge (Körper) A im E nennen wir bezüglich der Mr r parzellierbar, wenn sie sich als Vereinigungsmenge endlich vieler Parzellen darstellen läßt, die alle der Mr r assoziiert sind, wobei außerdem der Durchschnitt einer beliebigen Auswahl der verwendeten Parzellen wieder eine der Mr r assoziierte Parzelle ist. Für einen parzellierbaren Körper A gilt also eine Darstellung $A = P_1 \cup \ldots \cup P_n$, wobei $P_{\nu_1} \cap \ldots \cap P_{\nu_m}$ für jede Indizeskombination $\nu_1, \ldots, \nu_m$ und jedes $m = 1, \ldots, n$ eine der Mr r assoziierte Parzelle ist. Das Parzellensystem $\langle P_1, \ldots, P_n \rangle$ stellt eine der Mr r assoziierte Parzellierung (abgekürzt Pz) $\mathfrak{p} = \mathfrak{p}(A; r)$ von A dar. Die Anzahl n der verwendeten Parzellen nennen wir den Grad von $\mathfrak{p}$.

Ein Körper des Konvexrings, insbesondere ein Polyeder, ist bezüglich jeder Mr parzellierbar, da sich ein solcher als Vereinigungsmenge endlich vieler konvexer Körper darstellen läßt, und der Durchschnitt konvexer Körper stets wieder ein solcher ist; die leere Menge 0 wird hier auch als konvex angenommen.

6. Schnitteigenschaft der Parzellierungen. Ist der Körper A im E bezüglich der Mr r parzellierbar und ist E' ein der Mr r assoziierter Teilraum, so ist auch der Durchschnitt $A' = A \cap E'$ als Körper im E bezüglich r parzellierbar. Ist r' die beigeordnete Mr im E', die von den zu E' parallelen Richtungen

von r im E gebildet wird, so ist A' auch als Körper im E' bezüglich der Mr r' parzellierbar. Die beiden Aussagen bestätigen sich mit der folgenden Bemerkung: Ist $\langle P_1, \ldots, P_n \rangle$ eine der Mr r assoziierte Pz von A im E, so ist $\langle P_1 \cap E', \ldots, P_n \cap E' \rangle$ sowohl eine Pz von A' im E bezüglich der Mr r, als auch eine Pz von A' im E' bezüglich der Mr r'.

7. Normale Körper. Wir definieren: Ein normaler Körper des E ist eine kompakte Punktmenge A, die bezüglich jeder beliebig gewählten Mr r parzellierbar ist, wobei die Grade der Pz gleichmäßig beschränkt vorausgesetzt werden. Ist A ein normaler Körper, so gibt es also eine nur von A abhängige natürliche Zahl w derart, daß zu jeder Mr r eine Pz $p = p(A; r) = \langle P_1, \ldots, P_n \rangle$ aufgewiesen werden kann, so daß $n \leqq w$ ausfällt. Die Zahl w heißt Gradschranke. Die leere Menge 0 ist normal.

8. Schnitteigenschaften normaler Körper. Ist A ein normaler Körper im E und bezeichnet E' einen Teilraum von E, so ist auch der Durchschnitt $A' = A \cap E'$ als Punktmenge im E betrachtet ein normaler Körper. Weiter ist A' auch als Punktmenge im E' aufgefaßt ein normaler Körper. Bei beiden Interpretationen gilt für die Gradschranken die Ungleichung $w' \leqq w$. Unsere Behauptungen folgen unmittelbar aus der Definition der normalen Körper in Verbindung mit den Schnitteigenschaften der Pz.

9. Die Klasse der normalen Körper. Die Gesamtheit der normalen Körper im E fassen wir zur Körperklasse $\mathfrak{N}$ zusammen. $\mathfrak{N}$ ist bewegungsfrei, so daß aus $A \in \mathfrak{N}$, $A \simeq B$ (kongruent) auf $B \in \mathfrak{N}$ geschlossen werden kann. Ferner ist $\mathfrak{N}$ additiv, da aus $A, B \in \mathfrak{N}$, $A \cap B = 0$ auch $A \cup B \in \mathfrak{N}$ folgt. $\mathfrak{N}$ ist dagegen nicht ein Mengenring, da man aus $A, B \in \mathfrak{N}$ nicht ohne weiteres auf $A \cup B$, $A \cap B \in \mathfrak{N}$ schließen darf. Dagegen enthält $\mathfrak{N}$ wichtige Teilklassen mit der Ringeigenschaft, so den Konvexring und spezieller den Polyederring $\mathfrak{P}$. Daß die in Betracht gezogene Körperklasse $\mathfrak{N}$ tatsächlich alle Mengen und Figuren des Raumes umfaßt, die den Anspruch erheben können, als normal bezeichnet zu werden, kann hier nicht in theoretischer Weise näher begründet werden, sondern bleibt Sache der Erfahrung und des freien Ermessens.

10. Nulleinsfunktion. Aus formalen Gründen führen wir eine lediglich für Parzellen P definierte Hilfsfunktion ein. Die Nulleinsfunktion ε sei durch die Festsetzung $\varepsilon(P) = 1$ $(P \neq 0)$; $\varepsilon(P) = 0$ $(P = 0)$ erklärt.

11. Charakteristik einer Parzellierung. Es sei $p = p(A; r) = \langle P_1, \ldots, P_n \rangle$ eine Pz des normalen Körpers $A \in \mathfrak{N}$. Die durch den Ansatz

$$\chi(A, r, p) = \Sigma^1 \varepsilon(P_{\nu_1}) - \Sigma^2 \varepsilon(P_{\nu_1} \cap P_{\nu_2}) + \cdots \tag{1}$$

definierte, durch die der Mr r assoziierte Pz p von A eindeutig bestimmte ganze Zahl wollen wir Charakteristik χ der Pz p von A nennen; hierbei bezeichnet Σ^m eine Summation, die sich über alle Indizeskombinationen $\nu_1, \ldots, \nu_m$ der Indizesmenge $1, \ldots, n$ zur m-ten Klasse erstreckt.

12. Invarianzsatz. Ist $A \in \mathfrak{N}$ ein normaler Körper und bezeichnen $p = p(A; r)$ und $q = q(A; s)$ zwei den gleichen oder verschiedenen Mr r und s

assoziierte Pz von A, so gilt der Invarianzsatz

$$\chi(A, \mathfrak{p}, r) = \chi(A, \mathfrak{q}, s), \tag{2}$$

welcher also aussagt, daß die Charakteristik der Pz eines normalen Körpers von der gewählten Mr und von der individuellen Pz selbst unabhängig und demnach dem normalen Körper A invariant zugeordnet ist.

13. Beweis des Invarianzsatzes.

I. Es bezeichne $\mathfrak{R} = \mathfrak{R}(r)$ die Klasse der bezüglich der Mr r parzellierbaren Körper des E. Wir zeigen zunächst, daß sich über $\mathfrak{R}$ eine Funktion φ erklären läßt, welche die beiden nachfolgend formulierten Forderungen erfüllt:

1. Für eine Parzelle $P \in \mathfrak{R}$ gilt

$$\varphi(P) = \varepsilon(P) = 1 \quad (P \neq 0), \quad = 0 \quad (P = 0); \tag{3}$$

2. Falls $A, B, A \cup B, A \cap B \in \mathfrak{R}$ ist, so gilt

$$\varphi(A \cup B) + \varphi(A \cap B) = \varphi(A) + \varphi(B). \tag{4}$$

Falls eine solche Funktion φ existiert — was wir weiter unten nachweisen werden — so ist sie jedenfalls eindeutig bestimmt. Ist nämlich $\mathfrak{p} = \mathfrak{p}(A; r) = \langle P_1, \ldots, P_n \rangle$ eine Pz von A, so resultiert mit iterierter Anwendung von (3) und (4)

$$\varphi(A) = \chi(A, r, \mathfrak{p}), \tag{5}$$

womit gezeigt ist, daß jeder Funktionswert von φ eindeutig aus den beiden Eigenschaften abgeleitet werden kann. Andererseits ist mit (5) offensichtlich der Invarianzsatz im Sonderfall $s = r$ sichergestellt, da der Funktionswert links nicht von der individuell gewählten Pz $\mathfrak{p}$ abhängt.

Wir führen nun einen Existenzbeweis für φ durch vollständige Induktion nach der Dimensionszahl k.

A. Es sei $k = 1$. Es gibt lediglich zwei Mr r und $-r$ und es besteht $\mathfrak{R}(r) = \mathfrak{R}(-r)$ aus der leeren Menge 0 und aus denjenigen Punktmengen A, die sich aus endlich vielen, disjunkt liegenden abgeschlossenen Intervallen zusammensetzen. Bezeichnet $N(A)$ die Anzahl der Intervalle (Komponenten) von A, so ist mit

$$\varphi(A) = N(A) \tag{6}$$

bereits eine Funktion der verlangten Art gegeben. Man verifiziert mühelos, daß die Forderungen (3) und (4) erfüllt werden.

B. Es sei jetzt $k > 1$ und die Induktionsvoraussetzung sei erfüllt, wonach in jedem Raum kleinerer Dimension und für jede dort gewählte Mr eine Funktion der verlangten Art existiere.

Im Raum E wählen wir eine Mr r und eine assoziierte $(k-1)$-dimensionale Ebene E' mit der beigeordneten Mr r'. Ferner sei G eine auf E' orthogonal stehende Gerade. Den Punkt $G \cap E'$ wähle man als Ursprung eines in G liegenden Koordinatensystems mit der einen Koordinate x. Es sei weiter $E'[x]$

diejenige zu E' parallel liegende Ebene, welche G im Punkte mit der Koordinate x trifft. Wir betrachten nun einen Körper $A \in \mathfrak{R} = \mathfrak{R}(r)$ und eine der Mr r assoziierte Pz $p = p(A; r) = \langle P_1, \ldots, P_n \rangle$ von A. Bilden wir den Schnittkörper $A'[x] = A \cap E'[x]$, so gilt mit Rücksicht auf die Schnitteigenschaften der Pz sinngemäß $A'[x] \in \mathfrak{R}' = \mathfrak{R}'(r')$. Nach der Induktionsannahme existiert über $\mathfrak{R}'$ die Funktion φ' und wir setzen

$$f(A, x) = \varphi'(A'[x]). \tag{7}$$

Nach (5) gilt auch

$$f(A, x) = \chi(A'[x], r', p'), \tag{8}$$

wobei p' die von p in $E'[x]$ induzierte Pz von $A'[x]$ bedeutet. Mit (1) läßt sich auch

$$f(A, x) = \Sigma^1 \varepsilon(P_{\nu_1} \cap E'[x]) - \Sigma^2 \varepsilon(P_{\nu_1} \cap P_{\nu_2} \cap E'[x]) + \cdots \tag{9}$$

schreiben. Betrachten wir für den Augenblick einen rechts in (9) vorkommenden Summanden

$$\varepsilon(x) = \varepsilon(P_{\nu_1} \cap \cdots \cap P_{\nu_m} \cap E'[x]),$$

so bemerken wir, daß $\varepsilon(x)$ identisch verschwindet, wenn der eingetragene Parzellendurchschnitt leer ist und daß sich andernfalls zwei reelle Zahlen a und b, nämlich die beiden Stützkoordinaten des Parzellendurchschnitts ergeben, so daß $\varepsilon(x) = 1$ $(a \leqq x \leqq b)$, $= 0$ $(-\infty < x < a)$, $= 0$ $(b < x < \infty)$ ausfällt. Hierbei ist zu bedenken, daß der Parzellendurchschnitt selbst eine Parzelle ist und daß der Schnitt mit $E'[x]$ wieder eine Parzelle in $\mathfrak{R}'$ bleibt. Zusammengefaßt ergibt sich, daß $\varepsilon(x)$ eine ein- oder zweiwertige Stufenfunktion ist. Hieraus folgt, daß $f(A, x)$ als Ergebnis der Superposition von endlich vielen derartigen Funktionen eine endlichwertige Stufenfunktion sein muß. Nach diesem Befund ist es evident, daß der durch den Ansatz

$$f(A, x + 0) = \lim f(A, x + y) \quad [y > 0, y \to 0] \tag{10}$$

erklärte Grenzwert von rechts für jedes x existiert und lediglich an endlich vielen x-Stellen von $f(A, x)$ verschieden ausfallen kann. Mit

$$h(A, x) = f(A, x) - f(A, x + 0) \tag{11}$$

ist demnach eine Funktion gebildet, die nur an endlich vielen x-Stellen nicht verschwindet. Wir setzen

$$\varphi(A) = \sum_x h(A, x), \tag{12}$$

wobei formal über alle reellen x addiert werden soll, dagegen zu vermerken ist, daß die Summe nur endlich viele von Null verschiedenen Summanden aufweist.

Mit (12) ist über $\mathfrak{R}$ eine Funktion φ definiert. Wir zeigen, daß diese den Forderungen (3) und (4) genügt. In der Tat: Ist $P \in \mathfrak{R}$ eine nichtleere Parzelle

und sind a und b die Stützkoordinaten von P, so ist

$$f(P,x)=1 \quad (a\leqq x\leqq b), \quad =0 \quad (-\infty<x<a), \quad =0 \quad (b<x<\infty)$$

und $h(P,x)=1$ $(x=b)$, $=0$ $(x\neq b)$, so daß $\varphi(P)\doteq 1$ resultiert. Mit der trivialen Bemerkung, daß $\varphi(0)=0$ ist, ergibt sich (3). Gilt ferner $A, B, A\cup B, A\cap B\in\mathfrak{R}$, so folgt mit den Schnitteigenschaften der Pz analog

$$A'[x], B'[x], A'[x]\cup B'[x], A'[x]\cap B'[x]\in\mathfrak{R}'.$$

Nach Induktionsannahme, wonach φ' die Bedingung (4)′ erfüllt, folgt mit (7) zunächst

$$f(A\cup B,x)+f(A\cap B,x)=f(A,x)+f(B,x),$$

wo x auch durch $x+0$ ersetzt werden darf, so daß sich durch Subtraktion gemäß (11) weiter

$$h(A\cup B,x)+h(A\cap B,x)=h(A,x)+h(B,x)$$

und mit Summation über alle x schließlich

$$\varphi(A\cup B)+\varphi(A\cap B)=\varphi(A)+\varphi(B)$$

oder also (4) ergibt.

Wir bereits oben vermerkt, kann mit der nunmehr bewiesenen Existenz der Funktion φ auf

$$\chi(A,r,p)=\chi(A,r,q) \tag{13}$$

geschlossen werden.

II. Wir ziehen jetzt zwei verschiedene Mr r und s in Betracht. Für $A\in\mathfrak{R}$ gilt sowohl $A\in\mathfrak{R}(r)$, als auch $A\in\mathfrak{R}(s)$. Sind φ_r und φ_s die über $\mathfrak{R}(r)$ und $\mathfrak{R}(s)$ gemäß I. eindeutig konstruierten Funktionen, und sind p und q zwei den Mr r und s assoziierte Pz von A, so bestehen nach (5) die Beziehungen

$$\varphi_r(A)=\chi(A,r,p); \quad \varphi_s(A)=\chi(A,s,q). \tag{14}$$

1. Fall. Die beiden Mr r und s gestatten zwei total parallele assoziierte $(k-1)$-dimensionale Ebenen E_r' und E_s'. Bei der Konstruktion der Funktionen φ_r und φ_s kann ein- und dieselbe Gerade G als x-Achse dienen. Ist d der Abstand der Ebenen E_r' und E_s', so gelten für die nach den Ansätzen (7), (11) und (12) definierten Funktionen die Beziehungen

$$f_r(A,x)=f_s(A,x-d), \qquad h_r(A,x)=h_s(A,x-d)$$

und

$$\varphi_r(A)=\varphi_s(A). \tag{15}$$

2. Fall. Es gibt zwei den Mr r und s assoziierte nichtparallele $(k-1)$-dimensionale Ebene E_r' und E_s', die sich in einem $(k-2)$-dimensionalen Teilraum E'' schneiden. Es sei $t''=\langle t_3,\ldots,t_k\rangle$ eine Mr in E''. Bezeichnet t_1 bzw.

t_2 eine zu E'_r bzw. E'_s parallele und auf E'' orthogonal stehende Richtung, so stellt $t=\langle t_1, \ldots, t_k\rangle$ eine Mr im E dar, welche die Eigenschaft aufweist, daß sowohl E'_r als auch E'_s zu t assoziiert ist. Da A normal ist, gilt auch $A \in \mathfrak{R}(t)$. Bedeutet u eine der Mr t assoziierte Pz von A, so besteht für die über $\mathfrak{R}(t)$ eindeutig definierte Funktion φ_t nach (5) die Beziehung

$$\varphi_t(A) = \chi(A, t, u). \tag{16}$$

Da E'_r und E'_s der Mr t assoziiert sind, gilt nach der in I. durchgeführten Konstruktion

$$\varphi_r(A) = \chi(A, t, u); \qquad \varphi_s(A) = \chi(A, t, u) \tag{17}$$

und hieraus folgt

$$\varphi_r(A) = \varphi_s(A). \tag{18}$$

In beiden Fällen resultiert die nämliche, mit (15) und (18) ausgedrückte Gleichheit, die mit Rücksicht auf (14) gestattet, auf die Invarianzrelation

$$\chi(A, r, p) = \chi(A, s, q) \tag{19}$$

zu schließen. Damit ist der Invarianzsatz bewiesen.

14. Charakteristik normaler Körper. Durch den Ansatz

$$\chi(A) = \chi(A, r, p) \qquad [A \in \mathfrak{N}] \tag{20}$$

ist über der Klasse $\mathfrak{N}$ der normalen Körper des E eine ganzwertige Funktion, die Charakteristik χ eindeutig definiert, da die rechts stehende Charakteristik der der Mr r assoziierten Pz p nach dem soeben bewiesenen, mit (2) und (19) ausgedrückten Invarianzsatz weder von r noch von p abhängt.

15. Unabhängigkeit von der Dimensionszahl. Eine wichtige Eigenschaft der Charakteristik χ ist ihre Unabhängigkeit von der Dimension eines Einbettungsraumes. Liegt nämlich ein $A \in \mathfrak{N}$ ganz in einem i-dimensionalen Teilraum $E' = E_i$ von $E = E_k$, so läßt sich A auch als normaler Körper A' der sich auf den E' beziehenden Klasse $\mathfrak{N}'$ auffassen. Bedeutet χ' die über $\mathfrak{N}'$ gedeutete Charakteristik, so gilt

$$\chi'(A') = \chi(A). \tag{21}$$

Dies folgt unmittelbar aus der Tatsache, daß sich eine in E' gedeutete Pz p' von A' bezüglich der Mr r' auch als eine Pz p von A im E bezüglich der Mr r umdeuten läßt, wobei r aus r' durch Adjunktion von $k-i$ weiteren Richtungen des E erzeugt werden kann. Nach (20) folgert man sodann (21), wenn man bedenkt, daß die Charakteristiken der beiden Pz p und p' wegen (1) gleich ausfallen.

16. Schnittrekursion. Eine Gerade G des Raumes E sei als x-Achse gewählt. Es bezeichne $E'[x]$ die auf G orthogonal stehende $(k-1)$-dimensionale Ebene, welche G im Punkt mit der Koordinate x schneidet.

Die Charakteristik eines normalen Körpers A im E kann nach der bezüglich der Raumdimension rückgreifenden Schnittrekursion

$$\chi(A) = \sum_x \{\chi(A \cap E'[x]) - \chi(A \cap E'[x+0])\} \tag{22}$$

ermittelt werden. Hierbei bedeutet

$$\chi(A \cap E'[x+0]) = \lim \chi(A \cap E'[x+y]) \quad [y>0, y \to 0], \tag{23}$$

also den rechtsseitigen Grenzwert der Charakteristik des Schnittkörpers gebildet von A und $E'[x]$. Die in (22) vorgeschriebene Summation erstreckt sich über alle x, wobei zu beachten ist, daß der Summand lediglich für endlich viele Werte von x von Null verschieden ist.

Diese Schnittrekursion ergibt sich, wenn der Reihe nach die Beziehungen (20), (5), (11), (12) und auch (21) berücksichtigt werden. Sie nimmt bei der von uns vorgetragenen theoretischen Begründung der Charakteristik χ eine Schlüsselstellung ein; außerdem dient sie auch dazu, die Charakteristik eines vorgegebenen normalen Körpers praktisch zu ermitteln.

17. Additionstheorem. Sind mit A und B auch $A \cup B$ und $A \cap B$ normale Körper des E, so daß also $A, B, A \cup B, A \cap B \in \mathfrak{N}$ vorausgesetzt ist, so gilt das Additionstheorem

$$\chi(A \cup B) + \chi(A \cap B) = \chi(A) + \chi(B). \tag{24}$$

Dieses resultiert unmittelbar aus den Beziehungen (20), (5) und (4). Die Charakteristik χ ist als Funktional über $\mathfrak{N}$ additiv.

18. Multiplikationstheorem. Bezeichnet $A \times B$ das kartesische Produkt der beiden normalen Körper A und B, wobei vorausgesetzt wird, daß A und B in komplementären Teilräumen $E' = E_i$ und $E'' = E_{k-i}$ des E liegen, und ist auch $A \times B$ normal, gilt also $A, B, A \times B \in \mathfrak{N}$, so besteht das Multiplikationstheorem

$$\chi(A \times B) = \chi(A)\,\chi(B). \tag{25}$$

Beweis: Sind r' und r'' Mr in E' und E'', so bilden die in r' und r'' enthaltenen Richtungen insgesamt eine Mr r in E. Es sollen $\mathfrak{K}'$, $\mathfrak{K}''$ bzw. $\mathfrak{K}$ die Klassen der in E', E'' bzw. E bezüglich der Mr r', r'' bzw. r parzellierbaren Körper bezeichnen.

Zunächst zeigen wir, daß man aus $A \in \mathfrak{K}'$ und $B \in \mathfrak{K}''$ auf $A \times B \in \mathfrak{K}$ schließen kann. Sind $P \in \mathfrak{K}'$ und $Q \in \mathfrak{K}''$ zwei Parzellen, so folgt, daß $R = P \times Q$ eine zu $\mathfrak{K}$ gehörende Parzelle ist. Ist nämlich E_0 ein der Mr r assoziierter Teilraum, so gilt, wie leicht einzusehen, $E_0 = E_0' \times E_0''$, wo E_0' und E_0'' zwei den Mr r' und r'' assoziierte Teilräume von E' und E'' sind. Ferner gilt die elementar verifizierbare mengengeometrische Beziehung $R \cap E_0 = (P \cap E_0') \times (Q \cap E_0'')$, der man entnehmen kann, daß $R \cap E_0$ zusammenhängend ist, da das kartesische Produkt zusammenhängender Mengen selbst zusammenhängend ausfällt. Sind nun $p = p(A; r') = \langle P_1, \ldots, P_n \rangle$ und $q = q(B; r'') = \langle Q_1, \ldots, Q_m \rangle$ zwei Pz von

A und B, so ist zunächst leicht festzustellen, daß $C = A \times B$ als Vereinigung der nm Parzellen $R_{\nu\mu} = P_\nu \times Q_\mu$ ($\nu = 1, \ldots, n$; $\mu = 1, \ldots, m$) darstellbar ist. In der Tat stellt $\langle R_{11}, \ldots, R_{nm} \rangle$ sogar eine der Mr r assoziierte Parzellierung von C dar. Es muß gezeigt werden, daß der Durchschnitt $R_{\nu_1\mu_1} \cap \cdots \cap R_{\nu_l\mu_l}$ für jede Doppelindizeskombination selbst eine Parzelle ist. Dies folgt unmittelbar aus der mengengeometrischen Beziehung, wonach dieser Durchschnitt identisch ist mit

$$(P_{\nu_1} \cap \cdots \cap P_{\nu_l}) \times (Q_{\mu_1} \cap \cdots \cap Q_{\mu_l}).$$

Nun ist der Multiplikationssatz (25) zu begründen. Es bezeichne φ die in Teil I des Beweises des Invarianzsatzes über $\mathfrak{K}$ definierte Funktion. Setzen wir $\varphi(A \times B) = F(A; B)$, wobei wir uns A in $\mathfrak{K}'$ und B in $\mathfrak{K}''$ variiert denken, so gilt für zwei Parzellen $P \in \mathfrak{K}'$ und $Q \in \mathfrak{K}''$ zunächst (a) $F(P; Q) = \varepsilon(P \times Q) = \varepsilon(P)\,\varepsilon(Q)$. Mit Rücksicht auf die mengengeometrischen Beziehungen

$$(A_1 \cup A_2) \times B = (A_1 \times B) \cup (A_2 \times B); \quad (A_1 \cap A_2) \times B = (A_1 \times B) \cap (A_2 \times B)$$
$$A \times (B_1 \cup B_2) = (A \times B_1) \cup (A \times B_2); \quad A \cap (B_1 \times B_2) = (A \times B_1) \cap (A \times B_2)$$

lassen sich die Additionstheoreme (b)

$$F(A_1 \cup A_2; B) + F(A_1 \cap A_2; B) = F(A_1; B) + F(A_2; B)$$
$$F(A; B_1 \cup B_2) + F(A; B_1 \cap B_2) = F(A; B_1) + F(A; B_2)$$

folgern. Sind p und q die bereits oben in Betracht gezogenen Pz von A und B, so resultiert mit mehrfach iterierter Verwendung des Additionstheorems (b) die Summendarstellung (c)

$$F(A; B) = \sum_{i=1}^{n} \sum_{j=1}^{m} (-1)^{i+j}\, \Sigma^i\, \Sigma^j F(P_{\nu_1} \cap \cdots \cap P_{\nu_i}; Q_{\mu_1} \cap \cdots \cap Q_{\mu_j}),$$

wobei sich die rechte Seite mit Rücksicht auf (a) als Produkt

$$\left\{ \sum_{i=1}^{n} (-1)^i\, \Sigma^i\, \varepsilon(P_{\nu_1} \cap \cdots \cap P_{\nu_i}) \right\} \left\{ \sum_{j=1}^{m} (-1)^j\, \Sigma^j\, \varepsilon(Q_{\mu_1} \cap \cdots \cap Q_{\mu_j}) \right\}$$

schreiben läßt. Dieses erscheint nach (1) als das Charakteristikenprodukt $\chi(A, r', p)\, \chi(B, r'', q)$, so daß sich mit (5) $F(A; B) = \varphi(A \times B) = \varphi(A)\, \varphi(B)$ oder mit der Bemerkung, daß für normale Körper C nach (5) und (20) $\varphi(C) = \chi(C)$ gilt, schließlich $\chi(A \times B) = \chi(A)\, \chi(B)$ ergibt, was zu beweisen war.

19. Komplementärformel. Es bezeichne $A \in K$ einen normalen Körper, der in der Kugel K enthalten ist. Der bezüglich K komplementäre Körper A^* von A ist durch $A^* = K - A_0$ erklärt, wo A_0 den offenen Kern von A bedeutet. A_1 sei die abgeschlossene Hülle des offenen Kerns A_0 von A. Endlich sollen $\hat{A}$ und $\hat{A}_1$ die Ränder von A und A_1 bezeichnen. Wir setzen voraus, daß alle genannten, abgeschlossenen Mengen normal seien, so daß also $A, A^*, A_1, \hat{A}, \hat{A}_1 \in \mathfrak{K}$ gilt.

Zwischen den Charakteristiken der Körper A, A^* und der Randcharakteristik besteht dann die Komplementärformel

$$\chi(A^*) + \chi(A) = 1 + \chi(\hat{A}). \tag{26}$$

Beweis: Mit den Bemerkungen $\chi(K) = 1$ und

$$A^* \cup A_1 = K; \qquad A^* \cap A_1 = \hat{A}_1 \tag{a}$$

$$A_1 \cup \hat{A} = A; \qquad A_1 \cap \hat{A} = \hat{A}_1 \tag{b}$$

resultieren mit Anwendung des Additionstheorems (24) die Beziehungen

$$\chi(A^*) + \chi(A_1) = 1 + \chi(\hat{A}_1) \tag{a}$$

$$\chi(A_1) + \chi(\hat{A}) = \chi(A) + \chi(\hat{A}_1). \tag{b}$$

Durch Subtraktion ergibt sich die gewünschte Formel (26).

20. Nerv der Parzellierung und Charakteristik von EULER-POINCARÉ. Ist $\mathfrak{p} = \mathfrak{p}(A; r)$ eine der Mr r assoziierte Pz des normalen Körpers $A \in \mathfrak{R}$, so ist nach (20) die Charakteristik $\chi(A)$ identisch mit der nach Ansatz (1) eingeführten Charakteristik der Pz $\mathfrak{p}$. Diese wiederum ist durch die rein kombinatorischen Angaben über die vorkommenden leeren und nichtleeren Parzellendurchschnitte in $\mathfrak{p}$ eindeutig festgelegt.

Diesem Zusammenhang können wir formelmäßig Ausdruck geben, wenn wir die Anzahl $n_i (i = 1, \ldots, n)$ der vorkommenden Parzellen-i-tupel $\langle P_{\nu_1}, \ldots, P_{\nu_i} \rangle$ mit nichtleerem Durchschnitt $P_{\nu_1} \cap \cdots \cap P_{\nu_i} \neq 0$ einführen. Die Charakteristik von A kann dann als alternierende Summe dieser Anzahlen n_i bezüglich der Pz $\mathfrak{p}$ dargestellt werden, indem, wie mit (20) und (1) abgelesen werden kann,

$$\chi(A) = \Sigma_1^n (-1)^{i-1} n_i \tag{27}$$

gilt. Diese Formel bekommt dann noch eine besondere Deutung, wenn man der durch die Pz $\mathfrak{p}$ gegebenen Überdeckung von A den sog. Nerv zuordnet. Der Nerv ist ein euklidischer simplizialer Zellenkomplex im Sinne der kombinatorischen Topologie. In unserem Falle kann die Konstruktion des Nervs der Pz $\mathfrak{p}$ wie folgt durchgeführt werden: In einem euklidischen Hilfsraum ausreichend hoher Dimension ordne man den n Parzellen P_ν $(\nu = 1, \ldots, n)$ n Punkte p_ν $(\nu = 1, \ldots, n)$ eineindeutig zu und füge dem zu konstruierenden Zellenkomplex das durch die i Punkte $p_{\nu_1}, \ldots, p_{\nu_i}$ aufgespannte $(i-1)$-dimensionale Simplex genau dann zu, wenn die entsprechenden Parzellen einen nichtleeren Durchschnitt aufweisen. Der entstehende simpliziale Komplex ist der Nerv der Pz $\mathfrak{p}$. Die oben eingeführte Anzahl n_i ist offenbar die Anzahl der $(i-1)$-dimensionalen Simplexe des Nervs. Die in (27) rechts stehende alternierende Summe ist definitionsgemäß die Euler-Poincarésche Charakteristik des Nervs. Fassen wir die mit (27) ausgedrückte Tatsache in Worten anders zusammen, so läßt sich sagen: Die Charakteristik eines normalen Körpers ist identisch mit der Euler-Poincaréschen Charakteristik des Nervs einer seiner Parzellierungen.

21. Spezielle normale Körper und ihre Charakteristik. Nachfolgend sollen einige einfachste elementare normale Körper im E_k für $k \geqq 2$ und ihre Charakteristik angegeben werden. Der Nachweis der Normalität und die Ermittlung des χ-Wertes, die induktiv nach dem Additionssatz (24) oder nach der Schnittrekursion (22) eingeleitet werden kann, ist einfach und kann hier übergangen werden.

Bezeichnet K die k-dimensionale Vollkugel, so gilt $K \in \mathfrak{N}$ und

$$\chi(K) = 1. \tag{28}$$

Ist S die $(k-1)$-dimensionale Sphäre, also die Randfläche von K, so gilt $S \in \mathfrak{N}$ und

$$\chi(S) = 1 + (-1)^{k-1}. \tag{29}$$

Bedeutet K_h die Vollkugel K mit h sich gegenseitig nicht störenden zylindrischen Durchbohrungen, und S_h die Randfläche von K_h, so gilt K_h, $S_h \in \mathfrak{N}$ und es ist

$$\chi(K_h) = 1 + (-1)^k h, \tag{30}$$

$$\chi(S_h) = [1 + (-1)^{k-1}](1 - h). \tag{31}$$

Bedeutet K_l die Vollkugel K mit l sich gegenseitig nicht störenden kugelförmigen Ausbohrungen, so gilt $K_l \in \mathfrak{N}$ und

$$\chi(K_l) = 1 - (-1)^k l. \tag{32}$$

II. Teil

1. Integrale vom Croftonschen Typ. Es sei $A \in \mathfrak{N}$ ein normaler Körper und E' bezeichne eine im Raum $E = E_k$ frei bewegliche i-dimensionale Ebene $E' = E_i$ $(1 \leqq i \leqq k-1)$. Es soll nachgewiesen werden, daß das Integral vom Croftonschen Typ

$$\int \chi(A \cap E')\, dE' \tag{1}$$

als eigentliches Riemannsches Integral existiert, wobei dE' die kinematische Dichte von E' im Sinne der Integralgeometrie bedeutet.

Es gilt bekanntlich

$$dE' = d\overline{E}'\, d\overline{\overline{E}}', \tag{2}$$

wo $d\overline{E}'$ die Translationsdichte und $d\overline{\overline{E}}'$ die Drehdichte von E' darstellt.

Ist E'' die durch den festen Ursprung Z hindurchgehende, zu E' total orthogonale $(k-i)$-dimensionale Ebene, so ist $d\overline{E}' = d\mathfrak{p}$, wobei $\mathfrak{p} = E' \cap E''$ den Schnittpunkt von E' und E'' und $d\mathfrak{p}$ das Volumdifferential von $\mathfrak{p}$ in E'' bezeichnen. Wir betrachten zunächst das Translationsintegral

$$J(A, E'') = \int \chi(A \cap E')\, d\overline{E}'. \tag{3}$$

Die Existenz als eigentliches Riemannsches Integral ist leicht einzusehen, wenn man eine Pz $p = p(A; r) = \langle P_1, \ldots, P_n \rangle$ von A so einführt, daß die Ebene E' der Mr r assoziiert ist. Zerlegt man den Integranden $\chi(A \cap E')$ gemäß (1) I. Teil, so muß lediglich sichergestellt werden, daß

$$J(P, E'') = \int \chi(P \cap E')\, d\bar{E}' \tag{4}$$

für eine der Mr r assoziierte Parzelle P im genannten Sinne existiert. In der Tat ist die zu integrierende Funktion identisch mit der charakteristischen Funktion der in E'' liegenden Punktmenge $P|E''$, die durch orthogonale Projektion von P auf E'' erzeugt wird. $P|E''$ ist eine der dem E'' beigeordneten Mr r'' assoziierte Parzelle im E'', die im Jordanschen Sinne meßbar ist. So ergibt sich

$$J(P, E'') = \mu(P|E''), \tag{5}$$

wo μ das $(k-i)$-dimensionale Maß anzeigt, das in unserem Falle mit dem Jordanschen Inhalt zusammenfällt. Nach (3) läßt sich jetzt

$$J(A, E'') = \Sigma^1 \mu(P_{\nu_1}|E'') - \Sigma^2 \mu(P_{\nu_1} \cap P_{\nu_2}|E'') + \cdots \tag{6}$$

schreiben, wobei sich die Summation Σ^m über alle Indizeskombinationen $\nu_1, \ldots, \nu_m$ zur Klasse m von $1, \ldots, n$ erstreckt.

Es soll jetzt ferner gezeigt werden, daß $J(A, E'')$ als Funktion der um Z frei drehbaren E'' in einem zu präzisierenden Sinne stetig ist.

Wir wählen eine Vollkugel K mit Zentrum in Z so, daß sie den Körper A enthält; sie sei im folgenden festgedacht. Für zwei durch Z hindurchgehende $(k-i)$-dimensionale Ebenen E'' und E_0'' setzen wir $d(E'', E_0'') = \tilde{d}(E'' \cap K, E_0'' \cap K)$, wobei $\tilde{d}$ die im Raum der abgeschlossenen Punktmengen übliche Distanz bezeichnen soll. Durch d ist der Mannigfaltigkeit der durch Z hindurchgehenden E'' eine Metrik aufgeprägt. Nach geläufigen Schlüssen der Mengengeometrie ist für eine individuell gewählte abgeschlossene Punktmenge $M \subset K$ das Maß $\mu(M|E'')$ des Normalrisses $M|E''$ von M auf E'' eine stetige Funktion von E''; da der Raum der in K enthaltenen abgeschlossenen Mengen M kompakt ist, ist die Menge der Funktionen $\mu(M|E'')$, die sich für alle $M \subset K$ ergeben, gleichgradig stetig. Es gibt zu einem $\varepsilon > 0$ ein ϱ derart, daß simultan für alle $M \subset K$ $|\mu(M|E'') - \mu(M|E_0'')| < \varepsilon$ ausfällt, falls $d(E'', E_0'') < \varrho$ ist.

Wir ziehen jetzt zwei Ebenen E'' und E_0'' in Betracht, für welche $d(E'', E_0'') < \varrho$ gilt. Aus elementargeometrischen Gründen gibt es eine Mr r so, daß die beiden Ebenen E'' und E_0'' zu r assoziiert sind. Mit (6) folgt leicht $|J(A, E'') - J(A, E_0'')| < (2^w - 1)\,\varepsilon \leqq (2^w - 1)\,\varepsilon$, wobei w die Gradschranke des normalen Körpers A bedeutet. Damit ist die Stetigkeit von $J(A, E'')$ erwiesen. Mit Rücksicht auf (2) und (3) ist jetzt

$$\int \chi(A \cap E')\, dE' = \int J(A, E'')\, d\bar{E}' \tag{7}$$

als Integral über eine stetige Funktion im Riemannschen Sinne eigentlich.

Die mit (1) für $i=1,\dots,k-1$ angesetzten Integrale werden noch ergänzt durch

$$\int \chi(A\cap E_0)\,dE_0=\mu(A) \tag{8}$$

und

$$\int \chi(A\cap E_k)\,dE_k=\chi(A), \tag{9}$$

so daß die Croftonschen Integrale der Form (1) für alle $i=0,\dots,k$ erklärt sind.

2. Metrik im Raum der normalen Körper. Für $A, B\in\mathfrak{N}$ setzt man mit $i=0,\dots,k$:

$$d_i(A,B)=\frac{1}{c_i}\int|\chi(A\cap E_i)-\chi(B\cap E_i)|\,dE_i, \tag{10}$$

wobei

$$c_i=\binom{k}{i}\frac{\omega_{k-1}\cdots\omega_{k-i}}{\omega_1\cdots\omega_i}\qquad(i=1,\dots,k);\quad c_0=1 \tag{11}$$

und

$$\omega_m=\pi^{m/2}/\Gamma(1+m/2), \tag{12}$$

also gleich dem Inhalt der m-dimensionalen Einheitskugel ist.

d_i ist die i-te Partialmetrik in $\mathfrak{N}$. Die Totalmetrik ist durch

$$d(A,B)=\Sigma_0^k d_i(A,B) \tag{13}$$

definiert.

3. Konvergenz normaler Körper. Es sei $A\in\mathfrak{N}$ ein einzelner und $A_\nu\in\mathfrak{N}$ $(\nu=1,2,\dots)$ eine Folge normaler Körper. Diese konvergiert gegen A, geschrieben $A_\nu\to A$ $(\nu\to\infty)$, falls $d(A,A_\nu)\to 0$ $(\nu\to\infty)$.

4. Stetige Funktionale. Ein über $\mathfrak{N}$ definiertes Funktional Φ, das jedem $A\in\mathfrak{N}$ die reelle Zahl $\Phi(A)$ zuordnet, heißt stetig, wenn

$$\Phi(A_\nu)\to\Phi(A)\quad[A_\nu\to A,\ \nu\to\infty] \tag{14}$$

gilt.

5. Minkowskische Quermaßintegrale. Auf Grund der nachgewiesenen Existenz der Integrale vom Croftonschen Typ lassen sich über $\mathfrak{N}$ $k+1$ fundamentale Funktionale definieren, die als Verallgemeinerungen der von Minkowski für konvexe Körper begründeten Quermaßintegrale angesprochen werden dürfen. Diese bilden $k+1$ Maßzahlen, die einer über der Klasse $\mathfrak{N}$ der normalen Körper aufzubauenden Integralgeometrie als Basis zuzuweisen sind.

Das i-te Minkowskische Quermaßintegral eines normalen Körpers $A\in\mathfrak{N}$ definieren wir durch das Croftonsche Integral

$$W_i(A)=\frac{1}{c_i}\int\chi(A\cap E_i)\,dE_i\qquad(i=0,\dots,k), \tag{15}$$

wobei c_i die mit (11) festgelegten Konstanten bezeichnen. Es ist dann

(16) $$W_0(A) = V(A) = \text{Inhalt von } A$$

(17) $$kW_1(A) = F(A) = \text{Oberfläche von } A$$

(18) $$kW_2(A) = M(A) = \text{Integral der mittleren Krümmung von } A$$

. . . . . .

(19) $$kW_{k-1}(A) = N(A) = \text{Norm von } A$$

(20) $$(1/\omega_k)\,W_k(A) = \chi(A) = \text{Charakteristik von } A.$$

Diese Aufstellung läßt erkennen, daß die Minkowskischen Quermaßintegrale bis auf gewisse multiplikative Konstanten Maßzahlen darstellen, die zum Teil — wie die rechts angebrachten Bezeichnungen andeuten — bestens bekannt und auf andere Weise erklärt sind. In unserem Zusammenhang ist besonders hervorzuheben, daß durch die sich in natürlicher Weise aufdrängenden Croftonschen Integralbildungen die vollständige Skala dieser Maße einheitlich definiert ist, derart, daß die Maßzahlreihe mit dem Inhalt beginnt und mit der Charakteristik endigt. Die metrische Größe V und die topologische Invariante χ gehören also dem nämlichen grundlegenden Maßzahlsystem an. Die Darstellung (15) läßt weiter die enge Verflochtenheit zwischen der durch den gemeinsamen Integranden bedingten topologischen Ausmessung der Körper und der Bestimmung der metrischen Größen W_i klar hervortreten und belegt so die Wechselbeziehung, die zwischen Topologie und Integralgeometrie besteht.

6. Einfache Eigenschaften der Quermaßintegrale. Das i-te Minkowskische Quermaßintegral W_i ist ein über der Klasse der normalen Körper $\mathfrak{R}$ definiertes Funktional, welches bewegungsinvariant, definit und homogen vom Grad $k-i$ ist, so daß

(21) $$W_i(A) = W_i(B) \qquad [A \simeq B, \text{ kongruent}]$$

(22) $$W_i(A) \geqq 0$$

und

(23) $$W_i(\lambda A) = \lambda^{k-i} W_i(A) \qquad [\lambda \geqq 0]$$

gilt; λA bezeichnet den durch Dilatation mit λ aus A hervorgehenden homothetischen Körper.

Ferner ist W_i stetig und insbesondere gilt

(24) $$|W_i(A) - W_i(B)| \leqq d(A, B)$$

simultan für alle $i = 0, \ldots, k$, wo d die mit (13) eingeführte Distanz bezeichnet. Mit Respektierung der unserer Klasse $\mathfrak{R}$ aufgeprägten Totalmetrik gibt es demnach für keine der Maßzahlen W_i ein „Schwarzsches Phänomen".

7. Additionstheorem. Da der Integrand in der Croftonschen Darstellung (15) das Additionstheorem (24) des I. Teils befriedigt, überträgt sich die damit gegebene additive Eigenschaft auf die Minkowskischen Quermaßintegrale. Falls $A, B, A \cup B, A \cap B \in \mathfrak{N}$ vorausgesetzt wird, gilt die Beziehung

$$W_i(A \cup B) + W_i(A \cap B) = W_i(A) + W_i(B), \tag{25}$$

d. h. W_i ist additiv.

8. Quermaßintegrale von Kugel und Sphäre. Nachfolgend sollen die Maßzahlen für die im k-dimensionalen Raum E_k einbettbaren Kugeln und Sphären angegeben werden. Es bezeichne $K_p (0 \leqq p \leqq k)$ eine p-dimensionale Vollkugel vom Radius r und $S_q (0 \leqq q \leqq k-1)$ eine q-dimensionale Sphäre vom Radius r, also die Randfläche einer K_{q+1}. Für die Minkowskischen Quermaßintegrale $W_i (0 \leqq i \leqq k)$ dieser speziellen normalen Körper ergeben sich die folgenden Werte:

$$W_i(K_p) \begin{cases} = \dfrac{i!\,p!}{k!\,(i+p-k)!} \dfrac{\omega_i \omega_p}{\omega_{i+p-k}} r^{k-i} & [i+p \geqq k] \\ = 0 & [i+p < k], \end{cases} \tag{26}$$

$$W_i(S_q) \begin{cases} = \dfrac{2i!\,(q+1)!}{k!\,(i+q+1-k)!} \dfrac{\omega_i \omega_{q+1}}{\omega_{i+q+1-k}} r^{k-i} & [i+q \equiv k \pmod 2,\ i+q \geqq k] \\ = 0 & [i+q+1 \equiv k \pmod 2,\ i+q > k] \\ = 0 & [i+q < k]. \end{cases} \tag{27}$$

Das Resultat (26) ist geläufig, da K_p ein konvexer Körper ist. Zur Begründung von (27) kann man von den folgenden Bemerkungen ausgehen: Für $i+q<k$ ist $S_q \cap E_i$ fast immer leer. Für $i+q \geqq k$ ist $S_q \cap E_i$, falls nicht leer, eine $(i+q-k)$-dimensionale Sphäre. Für den Integranden $\chi(S_q \cap E_i)$ in der Darstellung (15) erhält man nach (29) I. Teil den Wert $1+(-1)^{i+q-k}$ oder anders ausgedrückt $2\chi(K_{q+1} \cap E_i)$ $[i+q \equiv k \pmod 2]$ oder 0 $[i+q+1 \equiv k \pmod 2]$, wobei K_{q+1} die von S_q berandete Vollkugel bezeichnet. So kann (27) auf (26) zurückgeführt werden.

9. Croftonsche Schnittformeln. Die innerhalb der Integralgeometrie der konvexen Körper bekannten Croftonschen Formeln können ohne weiteres auf normale Körper übertragen werden.

Bringen wir einen im E_k beweglichen i-dimensionalen Teilraum E_i mit dem Körper $A \in \mathfrak{N}$ zum Schnitt, so ist der Schnittkörper $A' = A \cap E_i$ ein i-dimensionaler normaler Körper $A' \in \mathfrak{N}'$. Der Beistrich $'$ soll hier wie nachfolgend andeuten, daß man sich auf den E_i bezieht. Bezeichnet W_j' $(j = 0, \ldots, i)$ das j-te Minkowskische Quermaßintegral im E_i, so ist $W_j'(A \cap E_i)$ als Funktion des beweglichen E_i im Riemannschen Sinne eigentlich integrierbar. Es gelten die Croftonschen Schnittformeln

$$\frac{1}{c_i} \int W_j'(A \cap E_i)\, dE_i = \frac{\omega_i \omega_{k-j}}{\omega_{k-i} \omega_{i-j}} W_j(A) \qquad [0 \leqq j \leqq i \leqq k-1]. \tag{28}$$

Die Hilfskonstanten c_i und ω_i sind mit (11) und (12) angegeben.

Beweis: Für den Integranden kann man nach (15) zunächst den Ausdruck

$$W_j'(A \cap E_i) = \frac{1}{c_j'} \int \chi(A \cap E_i \cap E_j')\, dE_j'$$

bilden, wobei E_j' einen im E_i beweglichen j-dimensionalen Teilraum und dE_j' die auf den E_i bezogene kinematische Dichte bezeichnen. Der Einsatz in (28) links liefert das Doppelintegral

$$\frac{1}{c_i' c_j'} \iint \chi(A \cap E_i \cap E_j')\, dE_j'\, dE_i.$$

Nun gilt die Dichterelation $dE_j'\, dE_i = dE_j\, dE_i''$, wo dE_i'' die Drehdichte des E_i um den Teilraum E_j im E_k bedeutet. Verwenden wir sie im oben stehenden Doppelintegral, so kann zunächst über die Drehung von E_i um den fest gedachten E_j integriert werden, wobei sich der Integrand nicht ändert, da ja $A \cap E_i \cap E_j = A \cap E_j$ bleibt. Indem nun

$$\int dE_i'' = \binom{k-j}{k-i} \frac{\omega_1 \cdots \omega_{k-j}}{\omega_1 \cdots \omega_{k-i} \omega_1 \cdots \omega_{i-j}}$$

gilt, erzielt man so mit einfachen Umrechnungen

$$\frac{\omega_i \omega_{k-j}}{\omega_{k-i} \omega_{i-j}} \frac{1}{c_j} \int \chi(A \cap E_j)\, dE_j$$

und somit nach (15) das gewünschte Ergebnis (28) rechts.

Bern (Schweiz), Hochfeldstr. 31

(Eingegangen am 2. Oktober 1958)

GERSTENHABER, M.
Math. Zeitschr. 71, 141/42 (1959)

Note on a Theorem of Wielandt

By

MURRAY GERSTENHABER *)

In a pape: in this Journal [6], H. WIELANDT determined all sets of $n \times n$ matrices with complex entries which are linear spaces over the real numbers and which have the property that every matrix in the space has only real proper values. Given integers $1 \leqq \alpha \leqq \beta \leqq n$, the set D of all pairs of indices (i, j) such that $\alpha \leqq i \leqq \beta$ and $\alpha \leqq j \leqq \beta$ will be called a diagonal block. If $A = (A_{ij})$ is any $n \times n$ matrix, the submatrix of those A_{ij} with (i, j) in D will be denoted by A_D. Let $D_1, \ldots, D_k$ be disjoint diagonal blocks whose union contains the diagonal and let $\mathfrak{H}(D_1, \ldots, D_k)$ denote the set of all matrices $A = (A_{ij})$ such that A_D is hermitian for $D = D_1, \ldots, D_k$ and such that $A_{ij} = 0$ if $j > i$ and $(i, j) \notin D_\alpha$ for any α. Then $\mathfrak{H}(D_1, \ldots, D_k)$ is a linear space over the real numbers of dimension n^2 the matrices in which have only real proper values, and WIELANDT proved that every linear space $\mathfrak{M}$ with these properties is in fact similar to some $\mathfrak{H}(D_1, \ldots, D_k)$. The proof WIELANDT gives of this remarkable result does not depend on any general algebraic theory, although one may observe as a consequence of his theorem that $\mathfrak{M}$ is in fact a Jordan algebra. In view of the existence of an extensive theory of Jordan algebras (developed principally by ALBERT, cf. [2]; for an extensive bibliography through 1955 see SCHAFER [5]) it is worthwhile to show that WIELANDT's theorem can also be obtained by specializing the general results.

Proof of WIELANDT's *theorem*. If $\mathfrak{M}$ is a real linear space of complex $n \times n$ matrices having only real proper values then, as observed by WIELANDT, the trace $Sp(A + tB)^2$ is non-negative for all A, B in $\mathfrak{M}$ and real t, whence $Sp\,AB$ is real. Fixing B, this is a linear condition on the coefficients of A. Therefore, $\dim \mathfrak{M} \leqq n^2$ and if $\dim \mathfrak{M} = n^2$ then $\mathfrak{M}$ is identical with the set of matrices A such that $Sp\,AB$ is real for all B in $\mathfrak{M}$. Since also $Sp(t_1 A_1 + t_2 A_2 + B)^3$ is real, it follows that $Sp(A_1 A_2 + A_2 A_1) B$ is real, therefore since $\mathfrak{M}$ has by assumption dimension n^2, $\mathfrak{M}$ is a Jordan algebra. The radical $\mathfrak{N}$ of $\mathfrak{M}$ is the set of all nilpotent elements of $\mathfrak{M}$, for these are the matrices A such that $Sp\,AB = 0$ for all B in $\mathfrak{M}$, and from this description obviously form an ideal. Let E denote the linear space on which the elements of $\mathfrak{M}$ operate. To prove the theorem of WIELANDT it is sufficient to show that if there exists no proper subspace R of E invariant under $\mathfrak{M}$ then $\mathfrak{M}$ is similar to the Jordan algebra $\mathfrak{H}$ of all $n \times n$ hermitian matrices. For we may assume the theorem true for all smaller values of n than that under consideration and the existence of such an R reduces the problem to cases of lower dimension. If $\mathfrak{M}$ has a non-zero

*) National Science Foundation Senior Postdoctoral Fellow 1958–1959.

radical $\mathfrak{N}$ then $E\mathfrak{N}=R$ is such a subspace, so we may in fact assume that $\mathfrak{M}$ is semi-simple and contains no nilpotent elements. It follows that $\mathfrak{M}$ is formally real, i.e., a sum of squares can not vanish, else there would exist a non-zero matrix A in $\mathfrak{M}$ for which $-A^2$ is also in $\mathfrak{M}$, but such an A would be nilpotent. Further, if the coefficient field of $\mathfrak{M}$ is extended to the complex numbers then one obtains the algebra Ω_n of all $n\times n$ complex matrices, else there would exist an non-zero matrix A in $\mathfrak{M}$ for which iA is also in $\mathfrak{M}$, which is impossible. Since Ω_n is simple (both as an associative and as a Jordan algebra) it follows that $\mathfrak{M}$ is simple, hence $\mathfrak{M}$ is a simple, formally real Jordan algebra of dimension n^2. All simple, formally real algebras were determined in the classical paper of JORDAN, VON NEUMANN, and WIGNER [4]: They are:

1. The set of all real symmetric $n\times n$ matrices.

2. The set $\mathfrak{H}_n$ of all $n\times n$ complex hermitian matrices.

3. The set of all $n\times n$ matrices (a_{ij}) with quaternion entries such that $a_{ij}=a_{ji}^*$, the quaternion conjugate.

4. An exceptional algebra of dimension 27 which ALBERT [1] has proved is not isomorphic to any Jordan algebra of linear transformations.

The dimensions of the algebras in 1, 2, and 3 are $\frac{1}{2}n(n+1)$, n^2, and $n(2n-1)$, respectively, hence comparing dimensions, $\mathfrak{M}$ is isomorphic to $\mathfrak{H}$. There exists therefore a linear mapping $\varphi:\mathfrak{H}\to\mathfrak{M}$ such that $\varphi(AB+BA)=\varphi(A)\,\varphi(B)+\varphi(B)\,\varphi(A)$, and extending as before the field of coefficients to the complex numbers, φ can be considered a mapping of Ω_n onto itself. Such a mapping has been called a semi-automorphism by ANCOCHEA [3] who proved that a semi-automorphism of a simple associative algebra is either an automorphism or an anti-automorphism. An automorphism of a central simple associative algebra being inner, it follows that there exists a matrix X in Ω_n such that either $\varphi(A)=X^{-1}AX$ for all A or $\varphi(A)=X^{-1}A'X$ for all A, where the prime denotes transposition. Since $\mathfrak{H}'=\mathfrak{H}$, it follows in either case that $\mathfrak{M}$ is similar to the set of all hermitian matrices. This ends the proof.

The author wishes to thank R. D. SCHAFER for valuable comments.

References

[1] ALBERT, A.: On a certain algebra of quantum mechanics. Ann. of Math. **35**, 65–73 (1934). — [2] ALBERT, A.: A structure theory for Jordan Algebras. Ann. of Math. **48**, 546–567 (1947). — [3] ANCOCHEA, G.: On semi-automorphisms of division algebras. Ann. of Math. **48**, 147–154 (1947). — [4] JORDAN, P., J. VON NEUMANN and E. WIGNER: On an algebraic generalization of the quantum mechanical formalism. Ann. of Math. **35**, 29–64 (1934). — [5] SCHAFER, R. D.: Structure and representation of non-associative algebras. Bull. Amer. Math. Soc. **61**, 469–484 (1955). — [6] WIELANDT, H.: Lineare Scharen von Matrizen mit reellen Eigenwerten. Math. Z. **53**, 219–225 (1950).

The Institute for Advanced Study, Princeton, N. J. and
The University of Pennsylvania, Philadelphia, Pa.

(Eingegangen am 16. Oktober 1958)

MacRobert, T. M.
Math. Zeitschr. 71, 143–145 (1959)

Infinite Series of *E*-Functions

By
T. M. MacRobert

§ 1. *Introductory.* In § 2 a number of series of E-functions are summed by expressing the E-functions as Barnes integrals and interchanging the order of integration and summation.

The Barnes integral employed is [*1*, p. 374]

$$E(p;\alpha_r:q;\varrho_s:z)=\frac{1}{2\pi i}\int\frac{\Gamma(\zeta)\prod_{r=1}^{p}\Gamma(\alpha_r-\zeta)}{\prod_{s=1}^{q}\Gamma(\varrho_s-\zeta)}z^{\zeta}\,d\zeta, \tag{1}$$

where $|\operatorname{amp} z|<\pi$ and the integral is taken up the η-axis, with loops, if necessary, to ensure that the origin lies to the left of the contour and the points $\alpha_1,\alpha_2,\ldots,\alpha_p$ to the right of the contour. Zero and negative integral values of the α's and ϱ's are excluded, and the α's must not differ by integral values. When $p<q+1$ the contour is bent to the left at each end.

Dougall's formula [*1*, p. 372]

$$\left\{F\left(\begin{matrix}\alpha, & \frac{1}{2}\alpha+1, & \beta, & \gamma, & \delta; & 1\\ & \frac{1}{2}\alpha, & \alpha-\beta+1, & \alpha-\gamma+1, & \alpha-\delta+1 & \end{matrix}\right)\right. \tag{2}$$
$$\left.=\frac{\Gamma(\alpha-\beta+1)\,\Gamma(\alpha-\gamma+1)\,\Gamma(\alpha-\delta+1)\,\Gamma(\alpha-\beta-\gamma-\delta+1)}{\Gamma(\alpha+1)\,\Gamma(\alpha-\beta-\gamma+1)\,\Gamma(\alpha-\gamma-\delta+1)\,\Gamma(\alpha-\delta-\beta+1)}\right\},$$

where $R(\alpha-\beta-\gamma-\delta)>-1$, will be required.

The Fourier series

$$\left(\sin\frac{1}{2}\vartheta\right)^{-2\zeta}=\frac{\Gamma(\frac{1}{2}-\zeta)}{\Gamma(\frac{1}{2})\,\Gamma(1-\zeta)}\left\{1+2\sum_{r=1}^{\infty}\frac{(\zeta;r)}{(1-\zeta;r)}\cos r\vartheta\right\}, \tag{3}$$

where $0\leqq\vartheta\leqq\pi$ will also be required. Here

$$(\zeta;r)=\zeta(\zeta+1)\ldots(\zeta+r-1),\quad r=1,2,3,\ldots;\qquad(\zeta;0)=1. \tag{4}$$

In proving (3) use is made of the formula [*1*, p. 340]

$$\int_0^{\pi/2}\cos(m-n)\vartheta\,(\cos\vartheta)^{m+n-2}\,d\vartheta=\frac{\pi\,\Gamma(m+n-1)}{2^{m+n-1}\,\Gamma(m)\,\Gamma(n)}, \tag{5}$$

where $R(m+n)>1$. If

$$\left(\sin\frac{1}{2}\vartheta\right)^{-2\zeta}=a_0+\sum_{r=0}^{\infty}a_r\cos r\vartheta,$$

$$a_0=\frac{1}{\pi}\int_0^{\pi}\left(\sin\frac{1}{2}\vartheta\right)^{-2\zeta}d\vartheta=\frac{2}{\pi}\int_0^{\pi/2}(\sin\varphi)^{-2\zeta}d\varphi$$

$$=\frac{2}{\pi}\int_0^{\pi/2}(\cos\varphi)^{-2\zeta}d\varphi=\frac{2\Gamma(1-2\zeta)}{2^{1-2\zeta}\Gamma(1-\zeta)\,\Gamma(1-\zeta)}=\frac{\Gamma(\frac{1}{2}-\zeta)}{\Gamma(\frac{1}{2})\,\Gamma(1-\zeta)}.$$

Also, if $r=1,2,3,\ldots$,

$$a_r=\frac{2}{\pi}\int_0^{\pi}\cos r\vartheta\left(\sin\frac{1}{2}\vartheta\right)^{-2\zeta}d\vartheta=\frac{4}{\pi}\int_0^{\pi/2}\cos(2r\varphi)\,(\sin\varphi)^{-2\zeta}d\varphi$$

$$=(-1)^r\frac{4}{\pi}\int_0^{\pi/2}\cos(2r\varphi)\,(\cos\varphi)^{-2\zeta}d\varphi=\frac{(-1)^r\Gamma(1-2\zeta)}{2^{-1-2\zeta}\Gamma(1-\zeta+r)\,\Gamma(1-\zeta-r)}$$

$$=\frac{2\Gamma(\frac{1}{2}-\zeta)\,(\zeta;r)}{\Gamma(\frac{1}{2})\,\Gamma(1-\zeta)\,(1-\zeta;r)};$$

and from this (3) follows.

§ 2. *Infinite Series.* The first summation is

$$(6)\qquad \sum_{n=0}^{\infty}\frac{(zt)^n}{n!}E\left(\begin{matrix}p;\alpha_r+n\\ q;\varrho_s+n\end{matrix}:z^{-2}\right)=E\left(\begin{matrix}p;\alpha_r\\ q;\varrho_s\end{matrix}:\frac{1}{z^2-tz}\right),$$

where $|\operatorname{amp} z|<\frac{1}{2}\pi$, $|\operatorname{amp}(z-t)|<\frac{1}{2}\pi$, $|t/z|<1$.

To prove it, substitute on the left from (1) and get

$$\sum_{n=0}^{\infty}\frac{(zt)^n}{n!}\frac{1}{2\pi i}\int\frac{\Gamma(\zeta)\prod\Gamma(\alpha_r+n-\zeta)}{\prod\Gamma(\varrho_s+n-\zeta)}z^{-2\zeta}d\zeta.$$

Here replace ζ by $\zeta+n$ and change the order of summation and integration. Then the series becomes

$$\frac{1}{2\pi i}\int\frac{\Gamma(\zeta)\prod\Gamma(\alpha_r-\zeta)}{\prod\Gamma(\varrho_s-\zeta)}z^{-2\zeta}F(\zeta;;t/z)\,d\zeta.$$

But

$$z^{-2\zeta}F(\zeta;;t/z)=z^{-2\zeta}(1-t/z)^{-\zeta}=(z^2-zt)^{-\zeta};$$

hence, using (1), the result follows.

The next summation is

$$(7)\qquad \left\{\begin{aligned}&\sum_{n=0}^{\infty}\frac{(l+2n)\,\Gamma(l+n)\,(l-\varrho_1+1;n)\,(k;n)\,\Gamma(\varrho_1-k)}{n!\,\Gamma(\varrho_1+n)\,\Gamma(l-k+1+n)\,z^n}\times\\ &\quad\times E\left(\begin{matrix}l-k+1+n,\alpha_1+n,\ldots,\alpha_p+n\\ l+2n+1,\ \varrho_1-k+n,\ \varrho_2+n,\ldots,\ \varrho_q+n\end{matrix}:z\right)=E(p;\alpha_r;q;\varrho_s:z),\end{aligned}\right.$$

where $|\operatorname{amp} z|<\pi$, $R(\varrho_1-k)>0$.

To prove it, apply (1) to the E-functions on the left, replace ζ by $\zeta+n$, and change the order of summation and integration, noting that

$$l+2n = l(\tfrac{1}{2}l+1;n)/(\tfrac{1}{2}l;n).$$

The expression then becomes

$$\frac{1}{2\pi i}\int \frac{\Gamma(\zeta)\,\Gamma(l-k+1-\zeta)\prod\Gamma(\alpha_r-\zeta)}{\Gamma(l+1-\zeta)\,\Gamma(\varrho_1-k-\zeta)\prod_{s=2}^{q}\Gamma(\varrho_s-\zeta)}\,z^{\zeta}\,\frac{\Gamma(l+1)\,\Gamma(\varrho_1-k)}{\Gamma(\varrho_1)\,\Gamma(l-k+1)}\times$$

$$\times F\left(\begin{matrix} l, & \tfrac{1}{2}l+1, & l-\varrho_1+1, & k, & \zeta; & 1 \\ \tfrac{1}{2}l, & & \varrho_1, & l-k+1, & l+1-\zeta & \end{matrix}\right)d\zeta.$$

On applying (2) to the last function it becomes

$$\frac{\Gamma(\varrho_1)\,\Gamma(l-k+1)\,\Gamma(l+1-\zeta)\,\Gamma(\varrho_1-k-\zeta)}{\Gamma(l+1)\,\Gamma(\varrho_1-k)\,\Gamma(l-k+1-\zeta)\,\Gamma(\varrho_1-\zeta)},$$

and, from (1), the result then follows.

Finally, if $0\leqq\vartheta\leqq\pi$, $|\operatorname{amp} z|<\pi$,

$$(8)\quad \left\{\begin{aligned} &E\left(\begin{matrix}\tfrac{1}{2},\alpha_1,\ldots,\alpha_p\\ 1,\varrho_1,\ldots,\varrho_q\end{matrix}:z\right)+2\sum_{r=1}^{\infty}\frac{\cos r\vartheta}{z^r}E\left(\begin{matrix}\tfrac{1}{2}+r,\ \alpha_1+r,\ldots,\alpha_p+r\\ 1+2r,\varrho_1+r,\ldots,\varrho_q+r\end{matrix}:z\right)\\ &\qquad\qquad=\sqrt{\pi}\,E\left(p;\alpha_r:q;\varrho_s:z/\sin^2\frac{1}{2}\vartheta\right).\end{aligned}\right.$$

For the expression on the left is equal to

$$\frac{1}{2\pi i}\int\frac{\Gamma(\zeta)\,\Gamma(\tfrac{1}{2}-\zeta)\prod\Gamma(\alpha_r-\zeta)}{\Gamma(1-\zeta)\prod\Gamma(\varrho_s-\zeta)}\,z^{\zeta}\times\left\{1+2\sum_{r=1}^{\infty}\frac{(\zeta:r)}{(1-\zeta:r)}\cos r\vartheta\right\}d\zeta,$$

and, on applying (3) and (1) the result follows.

Reference

[1] MacRobert, T. M.: Functions of a Complex Variable. London 1954.

Dept. of Math., The University, Glasgow, Scotland

(Eingegangen am 23. Oktober 1958)

BAKER, I. N.
Math. Zeitschr. 71, 146–153 (1959)

Fixpoints and Iterates of Entire Functions

By
I. N. BAKER

1. Introduction

In this paper $f(z)$ will always stand for an entire transcendental function of the complex variable z. For $n=1, 2, \ldots$ one may define the "natural iterates" $f_n(z)$ of $f(z)$ by

$$f_1(z)=f(z), \qquad f_n(z)=f_{n-1}(f(z))=f(f_{n-1}(z)). \tag{1}$$

These natural iterates are themselves entire and transcendental; for references to the literature see BAKER [1] and [2].

A fixpoint ξ of $f(z)$ is a zero of $f(z)-z$; the number $f'(\xi)$ is called the multiplier of ξ. More generally a fixpoint ξ of order n of $f(z)$ is a zero of $f_n(z)-z$ and we shall call $f_n'(\xi)$ the multiplier of ξ. A fixpoint ξ of order n is said to have exactly the order n when it is not a fixpoint of order less than n.

The fixpoints play a great role in the theory of natural iteration, both in the local theories of SCHROEDER, KOENIGS and others which deal with the behavior of the sequence $\{f_n(z)\}$ and the solution of various functional equations in the neighbourhood of fixpoints, and in FATOU'S [3] theory which treats the behavior of the sequence $\{f_n(z)\}$ in the large. Of the many problems about the fixpoints of $f(z)$ we shall take that of the existence and distribution of fixpoints of a given order. The results of [2] are carried further to prove Theorem 1 which asserts the existence of a great number (in a certain sense) of fixpoints of any order provided the order of $f(z)$ is less than $\frac{1}{2}$. Limitations of this result are discussed and in Theorems 2 and 3 applications to the distribution of values of the natural iterates are given.

2. Some results needed in the proof of Theorem 1

We shall use $M(f, r)$ to denote the maximum and $m(f, r)$ the minimum of $|f(z)|$ for $|z|=r$.

1. (BAKER [2, p. 124 Satz 1]). Let $f(z)$ be an entire function of order $\varrho<\frac{1}{2}$, $r(R)$ the smallest solution for given R of $m(f, r)=R$ and $M_{-1}(f, r)$ the inverse function of $M(f, r)$. Then to every $\delta>0$ there exists an $R_\delta>0$ so that for $R>R_\delta$ the inequalities

$$M_{-1}(f, R)\leqq r(R)\leqq\{M_{-1}(f, R)\}^{L(\varrho)+\delta} \tag{2}$$

hold, where

$$\frac{1}{L(\varrho)}=\operatorname*{Max}_{\varrho<\varrho'\leqq\frac{1}{2}}\left(1-\frac{\varrho}{\varrho'}\right)\cos(\pi\varrho').$$

2. (BAKER [2, p. 124 Hilfssatz 1]). Let $f(z)$ be an entire function and $M_{-1}(f, r)$ as in 1. For arbitrary $k>1$, $\delta>0$ there is an R_0 such that for all $R>R_0$ one has

$$M_{-1}(f, R^k)<(1+\delta)\{M_{-1}(f, R)\}^k<\{M_{-1}(f, R)\}^{k+\delta}. \tag{3}$$

3. (BAKER [2, p. 129 Satz 2]). Suppose one has numbers $B>A>0$ and $R>0$ and an entire function $g(z)$ such that for $|z|<A$ one has $|g(z)|<R$ while for $|z|=B$, $|g(z)|>R$ holds. Then there is in $A\leqq|z|<B$ a simple closed curve Γ which contains the origin and on which $|g(z)|=R$. In particular if $g(z)$ is a function of order less than $\frac{1}{2}$ there is an $R_0>0$ such that for all $R>R_0$ the annulus

$$M_{-1}(g, R)\leqq|z|<r(R+1) \tag{4}$$

contains a simple closed level curve $\Gamma: |g(z)|=R$ with the origin in its interior.

4. (BAKER [2, p. 131 formula (25)].) Let $\varepsilon>0$ be a given number and $f(z)$ an entire function of order $\varrho<\frac{1}{2}$; then there is a $c>1$ such that for all sufficiently great R the interval (R, R^c) contains an $\tilde{R}$ with

$$m(f, \tilde{R})>\{M(f, \tilde{R})\}^{-\varepsilon+\cos\pi\varrho}. \tag{5}$$

3. Existence of fixpoints of entire functions

FATOU [3] proved a result important for his theory of iteration in the large; this is that for a given $f(z)$ the set of all fixpoints (i.e. fixpoints of any order) has finite points of accumulation. Since the fixpoints of any finite order of $f(z)$ form a discrete set this means that any $f(z)$ has fixpoints of arbitrarily high exact order. When one turns to fixpoints of a given order it is clear that there need be none of order 1 as is shown by the example $f(z)=z+e^z$. ROSENBLOOM [6, 7], generalising an earlier result of FATOU [3] has shown that for any $f(z)$ the equation $f_n(z)-z=0$, $n>1$, has infinitely many solutions. His proof involves NEVANLINNA's second fundamental theorem of meromorphic functions. Although one knows that there are infinitely many solutions of $f_n(z)-z=0$ it is by no means proved that any of these is a fixpoint of the exact order n: it is thinkable that all these solutions are already fixpoints of order less than n, perhaps even of order 1. This is a difficulty which we do not overcome in our Theorem 1. However, by restricting the discussion to $f(z)$ of order less than $\frac{1}{2}$ we can use the essentially elementary methods of [2] to prove the following more quantitative results which have interesting applications in the theory of iteration (§ 4):

THEOREM 1. *Let the order of $f(z)$ be less than $\frac{1}{2}$. Let $n_l(r)$ be the number of solutions of $f_l(z)-z=0$ (counted according to multiplicity as zeros of $f_l(z)-z$) in $|z|\leqq r$; l is a positive integer. Write*

$$N_l(r)=\int_0^r \frac{n_l(t)-n_l(0)}{t}\,dt,$$

$$f(z)-z=c_p z^p+c_{p+1}z^{p+1}+\cdots, \qquad c_p\neq 0,$$

and let $k=k(l)$ be the constant whose existence will be proved in Lemma 1. Then for all sufficiently great r one has

$$\log\{M(f_l,r)-r^k\}<\log|c_p|+p\,k\log r+N_l(r^k). \tag{6}$$

We begin by proving the following

LEMMA 1. *If the order of $f(z)$ is $\varrho<\frac{1}{2}$, then to every $l=1,2,\ldots$ there is a constant $k=k(\varrho,l)$ and an R_0 such that for all $r>R_0$ the ring*

$$r\leqq|z|<r^k$$

contains a simple closed curve Γ which contains the origin and on which

$$|f_l(z)|=M(f_l,r)$$

holds.

PROOF OF LEMMA 1. Our proof is by induction on l. $l=1$: Put $M(f,r)=R$, $M_{-1}(f,R)=r$. Then from § 2 result 3, second part with $f(z)$ for $g(z)$, there exists an R_0 such that for all $R>R_0$ the annulus

$$r=M_{-1}(f,R)\leqq|z|<r(R+1)$$

contains a simple closed level curve Γ: $|f(z)|=R=M(f,r)$ with the origin in its interior. From § 2 (2) with $\delta=1$ we have

$$r(R)\leqq\{M_{-1}(f,R)\}^{L(\varrho)+1},$$

i.e.

$$\begin{aligned} r(R+1)&\leqq\{M_{-1}(f,R+1)\}^{L(\varrho)+1}\\ &<\{M_{-1}(f,R^2)\}^{L(\varrho)+1}\\ &<\{M_{-1}(f,R)\}^{3(L(\varrho)+1)}=r^{3(L(\varrho)+1)}\end{aligned}$$

for all sufficiently great R. Thus for $l=1$ we may take the k of Lemma 1 to be $3(L(\varrho)+1)$.

GENERAL INDUCTIVE STEP: Suppose the statement is true for a number l and let $k(l)$ be an admissible value of the constant k whose existence is asserted in the lemma. We shall prove the corresponding statement for $l+1$.

From § 2 result 4 with $\varepsilon=\frac{1}{2}\cos\pi\varrho$ follows the existence of a $c>1$ such that for all sufficiently great R there lies in (R,R^c) an $\tilde{R}$ with

$$m(f,\tilde{R})>\{M(f,\tilde{R})\}^{\frac{1}{2}\cos\pi\varrho}. \tag{7}$$

Now for all sufficiently great R (see § 2 result 2 with $k=c$, $\delta=1$)

$$M_{-1}(f_l,R^c)<\{M_{-1}(f_l,R)\}^{c+1},$$

and thus for all sufficiently great $r=M_{-1}(f_l,R)$ we have

$$M(f_l,r^{c+1})>\{M(f_l,r)\}^c=R^c.$$

Since $M(f_l, r)$ is a monotone and continuous function of r there is an $\tilde{r}$ with

$$r < \tilde{r} < r^{c+1} \tag{8}$$

and

$$M(f_l, \tilde{r}) = \tilde{R}, \tag{9}$$

where $\tilde{R}$ is the number in (7). Now by the induction hypothesis there is in

$$\tilde{r} \leqq |z| < \tilde{r}^{k(l)}$$

a simple closed curve Γ which contains the origin and on which

$$|f_l(z)| = M(f_l, \tilde{r}) = \tilde{R}. \tag{10}$$

Certainly Γ lies in

$$r \leqq |z| < r^{(c+1)k(l)}.$$

For z on Γ we have from (10), (7)

$$\begin{aligned} |f_{l+1}(z)| &= |f\{f_l(z)\}| \\ &\geqq m(f, \tilde{R}) > \{M(f, R)\}^{\frac{1}{2}\cos\pi\varrho} \\ &= [M\{f, M(f_l, r)\}]^{\frac{1}{2}\cos\pi\varrho} \\ &\geqq \{M(f_{l+1}, r)\}^{\frac{1}{2}\cos\pi\varrho}, \end{aligned}$$

and by an easy application of § 2 result 2 it is seen that this last quantity is greater than

$$M(f_{l+1}, r^{\frac{1}{2}\cos\pi\varrho})$$

for all sufficiently large r. Replacing $r^{\frac{1}{2}\cos\pi\varrho}$ by r we may assert that in the annulus

$$r^{\frac{3}{\cos\pi\varrho}} \leqq |z| < r^{\frac{3(c+1)k(l)}{\cos\pi\varrho}}$$

lies a simple closed curve Γ with the origin in its interior and on which

$$|f_{l+1}(z)| > M(f_{l+1}, r)$$

As in the proof of § 2 result 3 one deduces the existence of a simple closed curve Γ' (actually lying inside Γ) which has the origin in its interior, on which

$$|f_{l+1}(z)| = M(f_{l+1}, r)$$

and which lies in the ring

$$r \leqq |z| < r^{\frac{3(c+1)k(l)}{\cos\pi\varrho}}.$$

Thus one could take $k(l+1) = \dfrac{3(c+1)k(l)}{\cos\pi\varrho}$ and the inductive proof is complete.

Proof of Theorem 1: Let the solutions of $f_l(z) - z = 0$, other than $z = 0$ if this is a solution, be a_ν, $\nu = 1, 2, \ldots$. Take a fixed r so large that it is a valid r in Lemma 1 with our chosen value of l and also so that for no ν is a_ν

on the Γ curve (also of Lemma 1) which corresponds to this value r, and no a_ν has $|a_\nu| = r^{k(l)}$. (Since each a_ν lies on a Γ curve for at most one value of r this latter condition excludes at most a countable set of r-values.) For the a_ν which lie inside Γ, let $g(z, a_\nu)$ denote the GREEN's function of the interior of Γ with a_ν as fundamental singularity. If in the following Σ' of (11) the sum is extended over values of a_ν in the interior of Γ, each term occurring as many times as a_ν is a root of $f_l(z) - z = 0$ then the expression

$$\Phi(z) = \log|f_l(z) - z| + p\, g(z, o) + \Sigma' g(z, a_\nu) \tag{11}$$

is harmonic in and on Γ and so takes its minimum on Γ. Therefore we have for z in the interior of Γ

$$\Phi(z) > \log|M(f_l, r) - r^{k(l)}|$$

since Γ lies in the ring

$$r \leqq |z| < r^{k(l)}$$

while the GREEN's functions vanish on Γ. Suppose we enlarge Γ to the circle $z = r^{k(l)}$ and change each GREEN's function $g(z, o)$ or $g(z, a_\nu)$ in (11) to $g_1(z, o)$ or $g_1(z, a_\nu)$ resp., the GREEN's function for the circle $|z| < r^{k(l)}$, while extending the sum Σ' to include all a_ν with $|a_\nu| < r^{k(l)}$. Then the expression (11) becomes larger because the number of terms is increased while each GREEN's function becomes larger. We conclude

$$\log|f_l(z) - z| + p\, g_1(z, o) + \sum_{|a_\nu| < r^{k(l)}} g_1(z, a_\nu) > \log|M(f, r) - r^{k(l)}|. \tag{12}$$

Now

$$g_1(z, o) = \log \frac{r^{k(l)}}{|z|}$$

and

$$\log|f_l(z) - z| \sim \log|c_p| + p \log|z| \quad \text{for small } |z|.$$

Letting $z \to o$ in (12) gives for the left hand side

$$\log|c_p| + p\, k(l) \log r + \sum_{|a_\nu| < r^{k(l)}} g_1(o, a_\nu)$$

in which the last term may also be written $\sum g_1(a_\nu, o)$ or $\sum \log \frac{r^{k(l)}}{|a_\nu|}$ or $N_l(r^{k(l)})$.

This proves Theorem 1 [(5)] for all but a countable set of r-values and since all the quantities involved are continuous functions of r (at least for large r) the theorem holds in general. This proof follows closely the proof given by NEVANLINNA [4, p. 163–164] for the well-known Poisson-Jensen formula.

4. Applications of Theorem 1

Theorem 1 shows that $N_l(r) > \log M(f_l, r^{1/k}) - O(\log r)$. One sees that

$$\sum_{j=1}^{l-1} N_j(r) < \sum_{j=1}^{l-1} T(f_j, r) = \tau(r)$$

where T is the characteristic function of NEVANLINNA. An application of the results of PÓLYA [5] shows that $\tau(r)$ is of much smaller order than $\log M(f_l, r^{1/k})$. This does indeed allow us to conclude

$$N_l(r) \gg \sum_{j=1}^{l-1} N_j(r)$$

and *suggests* that there will be fixpoints of exact order l. Unfortunately there remains the possibility that all the solutions of $f_l(z)-z=0$ occur (with much lower multiplicities in some cases) among the solutions of $f_j(z)-z=0$ for $j=1, 2, \ldots, l-1$. One can, however, obtain some partial results of this kind, e.g.

THEOREM 2. *If the order of $f(z)$ is less than $\frac{1}{2}$ and no multiplier of a fixpoint of $f(z)$ of order 1 is 1 or a p-th root of unity for a prime p, then $f(z)$ has infinitely many fixpoints of the exact order p.*

PROOF: If a solution ξ of $f_p(z)-z=0$ is not a fixpoint of exact order p then it is a solution of $f(z)-z=0$, i.e. a fixpoint of order 1. Since $f'(\xi)\neq 1$ or a p-th root of unity we know that $f_p'(\xi)=\{f'(\xi)\}^p\neq 1$ and ξ is a single solution of $f_p(z)-z=0$. Suppose that there is but a finite number of fixpoints of exact order p, say m such points. Then $n_p(r)\leqq n_1(r)+m$ and

$$N_p(r) \leqq N_1(r) + O(\log r) < 2\,T(f, r) \tag{13}$$

for all sufficiently great r since

$$T(f, r) = O(\log r)$$

only for rational, i.e. non-transcendental functions and is a convex function of $\log r$. From Theorem 1

$$N_p(r) > \log M(f_p, r^{1/k}) - O(\log r)$$

and, using POLYA [5], there is a $c>0$ such that

$$\begin{aligned} M(f_p, r^{1/k}) &> M\left\{f, c\,M\left(f_{p-1}, \frac{r^{1/k}}{2}\right)\right\} \\ &> M(f, r^4) \\ &> M(f, r)^3 \end{aligned}$$

for all sufficiently great r. For these values then

$$\begin{aligned} N_p(r) &> 3\log M(f, r) - O(\log r) > 2\log M(f, r) \\ &\geqq 2\,T(f, r) \qquad \text{which contradicts (13).} \end{aligned}$$

We have excluded the multiplier 1 only for convenience. One sees that if

$$f(z) = \xi + (z-\xi) + c_h(z-\xi)^h + \cdots, \quad c_h \neq 0, \quad h \geqq 2$$

then

$$f_p(z) = \xi + (z-\xi) + p\,c_h(z-\xi)^h + \cdots.$$

Thus ξ has the same multiplicity as a solution of $f_p(z)-z=0$ as of $f(z)-z=0$. One may therefore include 1 among the admissible multipliers in Theorem 2 and has only to exclude the other p-th roots of unity. In particular one has the result:

THEOREM 2a. *If the order of $f(z)$ is less than $\frac{1}{2}$ and every multiplier of a fixpoint of $f(z)$ of order 1 is 1 then $f(z)$ has fixpoints of exact order p for every prime p.*

To form a function to which Theorem 2a applies one may take an arbitrary function $g(z)$ of order less than $\frac{1}{2}$ and define

$$f(z)=z+\{g(z)-z\}^2.$$

As a final application we note that not only the distribution of fixpoints but also the corresponding multipliers are subject to conditions. As an illustration we prove

THEOREM 3. *If each fixpoint of first order of $g(z)$ has a different multiplier then $g(z)$ is not the l-th iterate, $l>1$, of any entire function of order less than $\frac{1}{2}$.*

PROOF: Suppose $g(z)=f_l(z)$, $l>1$, with the order of $f(z)$ less than $\frac{1}{2}$. Then for at most one fixpoint ξ of $g(z)$ do we have

$$g'(\xi)=f_l'(\xi)=1,$$

i.e. the contribution of possible multiple roots of $f_l(z)-z=0$ to $N_l(r)$ is $O(\log r)$ and one easily sees that $f(z)$ has infinitely many fixpoints of exact order l.

But if η is any fixpoint $g(\eta)=\eta$ we have from $f(g(z))=g(f(z))$

$$f(\eta)=f(g(\eta))=g(f(\eta)) \tag{14}$$

and secondly

$$f'\{g(z)\}g'(z)=g'\{f(z)\}f'(z)$$

so that

$$f'(\eta)g'(\eta)=g'\{f(\eta)\}f'(\eta). \tag{15}$$

Now if $f'(\eta)=0$ then $g'(\eta)\neq 0$, i.e. this can be the case for at most one η. For all other η we have from (14) that $f(\eta)$ is a fixpoint of $g(z)$ while from (15) it has the same multiplier as does η. Therefore with one exception at most $f(\eta)=\eta$ and $f(z)$ has at most one fixpoint of exact order l, which gives a contradiction with the result proved above. The assumption $g(z)=f_l(z)$ must be false.

In particular one has the

COROLLARY: *If $g(z)$ is of finite order and all its fixpoints of order 1 have different multipliers then $g(z)$ is not the l-th iterate, $l>1$, of any entire function.*

PROOF: Suppose $g(z)=f_l(z)$, $l>1$, with $f(z)$ entire. Then it is proved in BAKER [2] that $f(z)$ is of order 0, so that Theorem 3 applies.

References

[1] Baker, I. N.: The iteration of entire transcendental functions and the solution of the functional equation $f\{f(z)\} = F(z)$. Math. Ann. **129**, 174–180 (1955). — [2] Baker, I. N.: Zusammensetzungen ganzer Funktionen. Math. Z. **69**, 121–163 (1958). — [3] Fatou, P.: Sur l'itération des fonctions transcendantes entières. Acta math. **47**, 337–370 (1926). — [4] Nevanlinna, R.: Eindeutige analytische Funktionen, 2. Aufl. Berlin: Springer 1953. — [5] Pólya, G.: On an integral function of an integral function. J. London Math. Soc. **1**, 12–15 (1926). — [6] Rosenbloom, P.: L'itération des fonctions entières. C. R. Acad. Sci. Paris **227**, 382–383 (1948). — [7] Rosenbloom, P.: The fixpoints of entire functions. Medd. Lunds Univ. mat. Sem., Suppl.-Band, M. Riesz 186–192 (1952).

Dept. of Math., University of Alberta, Edmonton, Alberta (Canada)

(Eingegangen am 1. November 1958)

AHRENS, J.
Math. Zeitschr. 71, 154–185 (1959)

Begründung der absoluten Geometrie des Raumes aus dem Spiegelungsbegriff*)

Von

JOACHIM AHRENS

Herrn FRIEDRICH BACHMANN zum 50. Geburtstag gewidmet

Einleitung

Klassische Untersuchungen von HESSENBERG und HJELMSLEV haben gezeigt, daß man sich bei einem axiomatischen Aufbau der absoluten Geometrie mit Vorteil des Rechnens mit Spiegelungen bedienen kann. In neueren Arbeiten von THOMSEN, ARNOLD SCHMIDT, BACHMANN, BAER, SPERNER, BERGAU, KLINGENBERG, KARZEL, LINGENBERG u.a. über ebene metrische Geometrie wird daher von vornherein ausgegangen von einer Gruppe und einem aus involutorischen Elementen bestehenden Erzeugendensystem dieser Gruppe, dessen Elemente gewisse Axiome erfüllen sollen.

Mit einem solchen Ansatz ist es gelungen (vgl. [1]), für einen Begriff von absoluter Geometrie der Ebene, in dem Forderungen über Anordenbarkeit oder freie Beweglichkeit nicht vorkommen, eine Begründung durchzuführen, d.h. die axiomatisch definierte Geometrie einer Behandlung mit den Methoden der analytischen Geometrie zugänglich zu machen.

Es ist das Ziel der vorliegenden Arbeit, für einen Begriff von absoluter Geometrie des Raumes, der von entsprechender Allgemeinheit ist, das gleiche zu leisten.

Wie in der Geometrie der Ebene gehen wir dabei aus von einer Gruppe $\mathfrak{G}$ und einem aus involutorischen Elementen bestehenden Erzeugendensystem $\mathfrak{S}$ dieser Gruppe, und verlangen, daß die Elemente von $\mathfrak{S}$ gewissen Axiomen genügen sollen. Die Axiome sind in den Bewegungsgruppen der klassischen euklidischen und nichteuklidischen dreidimensionalen Räume erfüllt, wenn für das Erzeugendensystem die Gesamtheit der Spiegelungen an den Ebenen des betreffenden Raumes genommen wird. Jedem solchen Paar $\mathfrak{G}$, $\mathfrak{S}$ ordnen wir dann einen „metrischen" Raum derart zu, daß die gegebene Gruppe $\mathfrak{G}$ als die Bewegungsgruppe dieses Raumes aufgefaßt werden kann, und ziehen zunächst aus dem Axiomensystem einige Folgerungen für den metrischen Raum (§ 1). Dem Lotensatz von HJELMSLEV geben wir eine räumliche Interpretation, mit deren Hilfe es gelingt, jedes Gruppenelement als Produkt von zwei involutorischen Elementen darzustellen (§ 2). Mittels eines Zusatzaxioms

*) Die vorliegende Arbeit wurde von der Deutschen Forschungsgemeinschaft gefördert. Ihr sage ich an dieser Stelle meinen Dank.

unterscheiden wir zwischen Räumen vom elliptischen und vom nichtelliptischen Typ. Im ersten Fall sind im metrischen Raum die projektiven Inzidenzaxiome erfüllt (§ 3). Weiter weisen wir nach, daß in unseren Bewegungsgruppen in mehrfacher Weise Bewegungsgruppen metrischer Ebenen im Sinne des Axiomensystems von BACHMANN [*1*] als Untergruppen enthalten sind, bzw., daß die Ebenen des metrischen Raumes als metrische Ebenen im Sinne von BACHMANN aufgefaßt werden können.

Für den Fall, daß der metrische Raum die projektiven Inzidenzaxiome nicht erfüllt, gewinnen wir durch eine Untersuchung der Relationen des Im-Büschel-Liegens und des Im-Bündel-Liegens von Ebenen (§ 5) den Anschluß an eine Arbeit von WINTERNITZ [*9*]. Hiermit ist, ohne Benutzung von Halbdrehungen, eine Erweiterung des gegebenen metrischen Raumes zu einem projektiven Idealraum durchgeführt, in welchem der Satz von PAPPUS-PASCAL und das Fano-Axiom gelten (§ 6). Im letzten Paragraphen (§ 7) nehmen wir eine erneute Gabelung unseres Aufbaus in einen „ordinären" Fall (nichteuklidische Metrik) und einen „singulären" Fall (euklidische Metrik) vor und weisen die Existenz einer projektiven Polarität im Idealraum (bzw. für den singulären Fall in einer Idealebene des Idealraumes) nach. Diese induziert im Idealraum eine Orthogonalität, welche für eigentliche Elemente die im metrischen Raum erklärte Orthogonalität ist. Daraus ergibt sich, daß jede Gruppe, welche dem Axiomensystem genügt, als Untergruppe einer Bewegungsgruppe eines projektiv-metrischen Raumes darstellbar ist. Der Anschluß an die Cayley-Kleinsche Auffassung der metrischen Geometrie ist damit erreicht und die axiomatisch gegebene räumliche Geometrie als algebraisierbar nachgewiesen.

§ 1. Bewegungsgruppen metrischer Räume

1. Das Axiomensystem.

Grundannahme. *Es sei eine Gruppe* $\mathfrak{G}$ *und ein (gegenüber inneren Automorphismen von* $\mathfrak{G}$*) invariantes, aus involutorischen Elementen bestehendes Erzeugendensystem* $\mathfrak{S}$ *von* $\mathfrak{G}$ *gegeben.*

Haben zwei involutorische Elemente Φ, Ψ aus $\mathfrak{G}$ ein involutorisches Produkt (d.h. gilt $\Phi\Psi = \Psi\Phi$, $\Phi \neq \Psi$), so schreiben wir dafür $\Phi \mid \Psi$. Stehen drei (bzw. vier) Erzeugende (Elemente aus $\mathfrak{S}$) α, β, γ (bzw. $\alpha, \beta, \gamma, \delta$) paarweise in dieser „Strichrelation", so schreiben wir auch $\alpha \mid \beta \mid \gamma$ (bzw. $\alpha \mid \beta \mid \gamma \mid \delta$).

Die Elemente aus $\mathfrak{S}$ werden auch *Ebenenspiegelungen* genannt und mit kleinen griechischen Buchstaben bezeichnet. Involutorische Elemente von $\mathfrak{G}$, die als ein Produkt $\alpha\beta$ zweier Erzeugender darstellbar sind, nennen wir *Geradenspiegelungen* und bezeichnen sie mit kleinen deutschen Buchstaben. Involutorische Elemente von $\mathfrak{G}$, die als ein Produkt $\alpha\beta\gamma$ mit $\alpha \mid \beta \mid \gamma$ darstellbar sind, werden *Punktspiegelungen* genannt und mit großen lateinischen Buchstaben bezeichnet. Das Axiomensystem handelt nur von Ebenen- und Punktspiegelungen:

Axiom 1. *Zu* P, Q, R *gibt es stets ein* ε *mit* $\varepsilon \mid P, Q, R$.
Axiom 2. *Aus* $P \neq Q$, $\alpha \neq \beta$; $P, Q \mid \alpha, \beta, \gamma$; $R \mid \alpha, \beta$ *folgt* $R \mid \gamma$.
Axiom 3. *Aus* $P \neq Q$; $\alpha, \beta, \gamma \nmid P, Q$ *folgt* $\alpha\beta\gamma \in \mathfrak{S}$.
Axiom 4. *Aus* $\pi \mid \vartheta$; $\alpha, \beta, \gamma \mid \pi, \vartheta$ *folgt* $\alpha\beta\gamma \in \mathfrak{S}$.
Axiom 5. *Gilt* $P \mid \alpha, \beta$, *so gibt es ein* Q *mit* $PQ \neq QP$ *und* $Q \mid \alpha, \beta$.
Axiom 6. *Zu jedem* ε *gibt es ein* P *mit* $P \mid \varepsilon$.

Wir nennen jedes Paar $\mathfrak{G}, \mathfrak{S}$, welches dem Axiomensystem genügt, eine *Bewegungsgruppe eines metrischen Raumes*. Im folgenden werden wir jedoch ein solches Paar $\mathfrak{G}, \mathfrak{S}$ einfach mit $\mathfrak{G}$ bezeichnen.

2. Der metrische Raum. Jeder Gruppe $\mathfrak{G}$, welche dem Axiomensystem genügt, wird nun ein *„metrischer Raum"* zugeordnet. Den Ebenenspiegelungen von $\mathfrak{G}$ ordnen wir *„Ebenen"*, den Punktspiegelungen von $\mathfrak{G}$ *„Punkte"* zu. Zwei Ebenen (bzw. zwei Punkte) seien genau dann einander gleich, wenn die entsprechenden Ebenenspiegelungen (bzw. Punktspiegelungen) gleich sind. Auf eine Unterscheidung in der Bezeichnungsweise können wir verzichten, Punkte seien mit denselben Buchstaben wie die zu ihnen gehörenden Punktspiegelungen, Ebenen mit denselben Buchstaben wie die zu ihnen gehörenden Ebenenspiegelungen bezeichnet. Der Leser kann es im folgenden jeweils leicht dem Zusammenhang entnehmen, ob gerade Elemente des metrischen Raumes oder die zugehörigen Spiegelungen aus $\mathfrak{G}$ gemeint sind.

Im klassischen elliptischen Raum bleibt bei der Spiegelung an einer Geraden auch die zu dieser polare Gerade fix, eine eineindeutige Zuordnung zwischen Geradenspiegelungen und Geraden ist also im allgemeinen ausgeschlossen. Wir geben daher eine andere Definition: Gilt $\alpha \mid \beta$, so sei durch $g \equiv (\alpha\beta)$ im metrischen Raum eine *„Gerade"* dargestellt. Zwei Geraden $(\alpha\beta)$ und $(\gamma\delta)$ seien genau dann identisch, wenn erstens $\alpha\beta = \gamma\delta$ (in $\mathfrak{G}$) gilt, und wenn zweitens für beliebige Erzeugende σ die Aussage $\sigma \mid \alpha, \beta$ äquivalent zu $\sigma \mid \gamma, \delta$ ist. Gilt dagegen zwar $\alpha\beta = \gamma\delta$, aber stimmt die Menge der mit α und β in Strichrelation stehenden Erzeugenden nicht mit der entsprechenden Menge für γ und δ überein, so sollen $(\alpha\beta)$ und $(\gamma\delta)$ zueinander *polar* heißen. Für die Bezeichnung von Geraden verwenden wir kleine lateinische Buchstaben, zugehörige Geradenspiegelungen werden mit den entsprechenden kleinen deutschen Buchstaben geschrieben.

Die Polarität im elliptischen Raum bewirkt ferner, daß dort eine Punktspiegelung einer Ebenenspiegelung gleich sein kann. Gilt $P = \varepsilon$, so nennen wir den Punkt P den *Pol* von ε und die Ebene ε die *Polarebene* von P. Auch eine Gleichung der Form $P = \alpha\beta\gamma$ kann bestehen, ohne daß $\alpha \mid \beta \mid \gamma$ gilt; wir schreiben aber im folgenden $P \equiv (\alpha\beta\gamma)$, wenn wir betonen wollen, daß der Punkt P durch α, β, γ „dargestellt" wird, d.h., wenn $P = \alpha\beta\gamma$ und $\alpha \mid \beta \mid \gamma$ ist (hierzu ist äquivalent, daß $P = \alpha\beta\gamma$ und $P \mid \alpha, \beta, \gamma$ gilt).

Es werden nun die Begriffe *Inzidenz* (I) und *Orthogonalität* ($\perp$) erklärt als symmetrische Relationen, welche zwischen Elementen des metrischen Raumes bestehen können:

Definition. *Es sei genau dann*
$\varepsilon \perp \delta$ *und* $\delta \perp \varepsilon$, *wenn (für die zugehörigen Spiegelungen)* $\varepsilon | \delta$ *gilt*;
$P I \varepsilon$ *und* $\varepsilon I P$, *wenn (für die zugehörigen Spiegelungen)* $P | \varepsilon$ *gilt*;
$g I \varepsilon$ *und* $\varepsilon I g$, *wenn* g *in der Form* $g \equiv (\varepsilon \delta)$ *darstellbar ist*;
$g \perp \varepsilon$ *und* $\varepsilon \perp g$, *wenn* g *in der Form* $g \equiv (\alpha\beta)$ *mit* $\alpha, \beta | \varepsilon$ *darstellbar ist*;
$P I g$ *und* $P I g$, *wenn es* α, β, γ *gibt mit* $g \equiv (\alpha\beta)$ *und* $P \equiv (\alpha\beta\gamma)$;
$h \perp g$ *und* $g \perp h$, *wenn es* α, β, γ *gibt mit* $g \equiv (\alpha\beta)$, $h \equiv (\alpha\gamma)$ *und* $\beta | \gamma$.

Innere Automorphismen von $\mathfrak{G}$ führen nach der Grundannahme Elemente des Erzeugendensystems $\mathfrak{E}$ wieder in Erzeugende über. Aus $\alpha | \beta$ folgt für die transformierten Elemente $\alpha^\Phi = \Phi^{-1}\alpha\Phi$ und $\beta^\Phi = \Phi^{-1}\beta\Phi$, daß $\alpha^\Phi \beta^\Phi = (\alpha\beta)^\Phi$ involutorisch ist, also daß auch $\alpha^\Phi | \beta^\Phi$ gilt. Aus $P | \varepsilon$ ergibt sich ebenso $P^\Phi | \varepsilon^\Phi$, und man sieht damit bei einem Vergleich mit den Definitionen leicht ein, daß durch einen inneren Automorphismus von $\mathfrak{G}$ im metrischen Raum eine eineindeutige Abbildung der Mengen der Ebenen, Geraden und Punkte je auf sich induziert wird, bei der Inzidenz und Senkrechtstehen erhalten bleiben. Wir nennen eine solche Abbildung eine *Bewegung des metrischen Raumes*. Ist das transformierende Element Φ dabei eine Erzeugende ε, so heiße die induzierte Bewegung des metrischen Raumes die Spiegelung Σ_ε an der Ebene ε. Wir werden später sehen (Satz 31), daß die Gruppe der inneren Automorphismen von $\mathfrak{G}$ isomorph zu $\mathfrak{G}$ selbst ist. Damit ist dann die Benennung von $\mathfrak{G}$ als Bewegungsgruppe eines metrischen Raumes gerechtfertigt.

Stehen Elemente des metrischen Raumes in der Relation der Inzidenz (I), so werden wir an Stelle von „inzidiert mit" auch Redeweisen wie „liegt auf", „geht durch" usw. verwenden. Desgleichen sagen wir anstatt „ist orthogonal zu" auch „steht senkrecht", „ist ein Lot (eine Lotebene) auf" und dergleichen.

Die Axiome 1 und 2 führen im metrischen Raum auf Existenz und Eindeutigkeit von Verbindungsebenen dreier Punkte (vgl. Satz 12); Axiom 2 ist das „Inzidenzaxiom" bei WINTERNITZ [9] (vgl. Satz 33, Axiom a).

Die Axiome 3 und 4 sind zwei Formen des Satzes von den drei Spiegelungen (vgl. hierzu Satz 22).

Axiom 5 besagt, daß zwei Ebenen, die mit einem Punkt P inzidieren, noch mindestens einen weiteren Punkt Q gemein haben (Q soll nicht auf einer eventuell existierenden Polarebene von P liegen). Dieses Axiom bringt zum Ausdruck, daß der metrische Raum höchstens dreidimensional sein soll (vgl. Axiom I_7 bei HILBERT [3]).

Axiom 6 schließlich fordert für den metrischen Raum, daß auf jeder Ebene mindestens ein Punkt liegt.

Sätze zitieren wir im folgenden durch ihre Nummer in runden Klammern, bei Axiomen setzen wir ein A, bei Hilfssätzen ein H vor die Nummer. Zahlen in eckigen Klammern beziehen sich stets auf das Literaturverzeichnis.

3. Erste Folgerungen aus dem Axiomensystem.

Satz 1 *(Polartetraeder). Aus* $\alpha | \beta | \gamma | \delta$ *folgt* $\alpha\beta\gamma\delta = 1$.

Beweis. Da $\alpha, \beta, \gamma, \delta$ paarweise kommutieren, so sind $\alpha\beta\gamma\delta$, $\alpha\beta\gamma$, $\alpha\beta\delta$, $\alpha\gamma\delta$, $\beta\gamma\delta$ involutorisch oder gleich 1. Von den Produkten zu je dreien ist

höchstens eines die Identität (aus $\alpha\beta\gamma = 1 = \alpha\beta\delta$ folgte $\gamma = \delta$), die anderen sind involutorisch, also Punktspiegelungen. Sei etwa $\beta\gamma\delta = A$, $\alpha\gamma\delta = B$, $\alpha\beta\delta = C$. Wäre nun $\alpha\beta\gamma\delta$ involutorisch, so folgte $\alpha, \beta, \delta \mid A, B$ und damit $\alpha\beta\delta = \varepsilon$ (A3)*). Aus $A, B \mid \alpha, \beta, \varepsilon$; $C \mid \alpha, \beta$; $A \neq B$; $\alpha \neq \beta$ ergäbe sich (A2) $\varepsilon \mid C$ im Widerspruch zu $\varepsilon = C$.

Als Ergänzung zu den Axiomen 3 und 4 beweisen wir den

Satz 2. *a) Aus* $P \neq Q$; $\alpha, \beta, \gamma \mid P, Q$; $\alpha\beta\gamma = \delta$ *folgt* $\delta \mid P, Q$.

b) Aus $\pi \neq \vartheta$; $\alpha, \beta, \gamma \mid \pi, \vartheta$; $\alpha\beta\gamma = \delta$ *folgt* $\delta \mid \pi, \vartheta$.

Beweis. a) Ist $\alpha = \beta$, so ist $\gamma = \delta$, also $\delta \mid P, Q$. Es sei also $\alpha \neq \beta$. Da P, Q mit α, β, γ kommutieren, gilt $\delta P = P\delta$, $\delta Q = Q\delta$. Angenommen, es ist nicht $\delta \mid P, Q$. Dann folgt $\delta = P$ oder $\delta = Q$. Sei etwa $\delta = P$. Nach (A5) gibt es ein R mit $R \mid \alpha, \beta$; $RP \neq PR$. (A2) ergibt $R \mid \gamma$, und damit $RP = R\alpha\beta\gamma = \alpha\beta\gamma R = PR$ (Widerspruch).

b) Analog zu a) gilt zunächst $\delta\pi = \pi\delta$, $\delta\vartheta = \vartheta\delta$. Wäre etwa $\pi = \delta$, so folgte $\alpha \mid \beta \mid \gamma$ und damit $\alpha \mid \beta \mid \gamma \mid \pi$ sowie $\alpha \mid \beta \mid \gamma \mid \vartheta$, also (1) $\alpha\beta\gamma\pi = 1 = \alpha\beta\gamma\vartheta$, $\pi = \vartheta$ (Widerspruch).

Über die Inzidenz von Punkt und Ebene gilt

Satz 3. *Gilt* $P \mid \alpha$, *so gibt es* β, γ *derart, daß* $P \equiv (\alpha\beta\gamma)$ *ist.*

Beweis. Es sei $P \equiv (\varrho\sigma\tau)$. Wir wählen (A5) Q und R mit $Q, R \neq P$; $Q \mid \varrho, \sigma$; $R \mid \tau, \alpha$. Nach (A1) gibt es ein ε mit $\varepsilon \mid P, Q, R$. Nun ist $P\alpha = (\varrho\sigma\varepsilon)(\varepsilon\tau\alpha) = \beta\gamma$ (A3). Aus $P \mid \alpha$ folgt $\gamma \mid \beta$. Nach (2a) gilt $\gamma, \beta \mid P$ und damit $\alpha \mid \beta, \gamma$; d.h. $P \equiv (\alpha\beta\gamma)$.

Zur Vorbereitung der Einführung von Verbindungsgeraden dient der

Satz 4. *Zu* P, Q, α *mit* $P, Q \mid \alpha$ *gibt es* β, π *mit* $\beta \mid P, Q, \alpha$; $P \equiv (\alpha\beta\pi)$.

(*) *Aus* $P \neq Q$; $P, Q \mid \alpha, \beta, \gamma$; $\beta, \gamma \mid \alpha$ *folgt* $\beta = \gamma$.

Beweis. Nach (3) ist P als $(\alpha\sigma\tau)$ darstellbar. Wir wählen (A5) ein R mit $R \neq P$, $R \mid \sigma, \tau$ und ein β mit $\beta \mid P, Q, R$ (A1). Dann gilt $\beta\sigma\tau = \pi$ (A3) und $\pi \mid P$ (2). Damit wird $P = \alpha\beta\pi$, und aus $P \mid \alpha, \beta, \pi$ folgt $\alpha \mid \beta \mid \pi$.

(*) Wie soeben bewiesen, gibt es ein δ mit $\delta \mid P, Q$; $P = (\alpha\delta\pi)$. Es ist $\pi\beta = \delta\alpha P\beta = \delta\alpha\beta P$. Aus $\delta, \alpha, \beta \mid P, Q$ ergibt sich (A3) $\delta\alpha\beta = \varepsilon$, $\varepsilon \mid P$ (2) und damit: $\pi\beta = \varepsilon P$ ist involutorisch. Nun ist $\beta\delta = \beta\alpha P\pi = \alpha P\pi\beta = \delta\beta$. Wäre $\delta \mid \beta$, so hätten wir $\beta \mid \alpha \mid \delta \mid \pi$, also (1) $\beta\alpha\delta\pi = 1$, $P = \beta$ im Widerspruch zu $P \mid \beta$. Daher gilt $\beta = \delta$, und ebenso schließt man auf $\gamma = \delta$.

Mit dem folgenden Satz ergänzen wir Satz 3:

Satz 5. *a) Gilt* $P \mid \alpha, \beta$ *und* $\alpha \mid \beta$, *so gibt es ein* γ *derart, daß* $P \equiv (\alpha\beta\gamma)$ *ist.*

b) Gilt $P \mid \alpha, \beta, \gamma$ *und* $\alpha \mid \beta \mid \gamma$, *so folgt* $P \equiv (\alpha\beta\gamma)$.

Beweis. a) Nach (A5) gibt es ein Q mit $Q \neq P$; $Q \mid \alpha, \beta$. Zu P, Q, α existiert (4) ein β', so daß $\beta' \mid P, Q, \alpha$; $P \equiv (\alpha\beta'\gamma)$ gilt. (4*) ergibt aber $\beta' = \beta$.

*) Eigentlich müßte es heißen: Nach (A3) *gibt es* ein ε, derart, daß $\alpha\beta\delta = \varepsilon$ ist. Aus Raumgründen wollen wir uns jedoch in Beweisen mit der obigen, etwas unpräzisen Ausdrucksweise begnügen.

b) Nach a) ist P darstellbar als $P \equiv (\alpha\beta\gamma')$. Es folgt $\gamma\gamma' = \gamma\beta\alpha P = \beta\alpha P\gamma = \gamma'\gamma$. Wäre nun $\gamma' \mid \gamma$, so hätten wir $\alpha \mid \beta \mid \gamma' \mid \gamma$, also (1) $P = \alpha\beta\gamma' = \gamma$ im Widerspruch zu $P \mid \gamma$. Daher gilt $\gamma' = \gamma$, $P = (\alpha\beta\gamma)$.

Satz (5a) besagt in der Sprechweise des metrischen Raumes: Aus $P\,\mathrm{I}\,\alpha, \beta$; $g \equiv (\alpha\beta)$ folgt $P\,\mathrm{I}\,g$. Satz 5b besagt, daß der Schnittpunkt dreier paarweise zueinander senkrechter Ebenen eindeutig bestimmt ist. Daß ein solcher Schnittpunkt stets existiert, zeigen wir in Satz 7.

Satz 6. *Zu jedem g gibt es drei verschiedene Punkte P, Q, R, so daß $P, Q, R\,\mathrm{I}\,g$ gilt.*

Beweis. Es sei $g \equiv (\alpha\beta)$. Zu α gibt es (A6) ein P mit $P \mid \alpha$.

1. Fall: $P \mid \beta$. Dann folgt $P\,\mathrm{I}\,g$ nach (5a).

2. Fall: $P = \beta$. P sei nach (3) dargestellt als $P \equiv (\alpha\sigma\tau)$. Aus $\alpha \mid \sigma \mid \tau$, $\alpha\sigma\tau = \beta$ folgt $\beta \mid \sigma$. Da $\alpha\beta\sigma = \tau$ involutorisch ist, stellt $(\alpha\beta\sigma)$ einen Punkt dar, welcher mit g inzidiert.

3. Fall: $P^\beta \neq P$. Nach (4) gibt es ein γ mit $\gamma \mid \alpha, P, P^\beta$. Durch Transformieren erhält man: $\gamma^\beta \mid \alpha^\beta = \alpha, P^\beta, P^{\beta\beta} = P$. Mit $P, P^\beta \mid \alpha$ folgt (4*) $\gamma^\beta = \gamma$. $\gamma = \beta$ ist wegen $P^\gamma = P$ ausgeschlossen, damit gilt $\alpha \mid \beta \mid \gamma$. Wäre $\alpha\beta\gamma = 1$, so wäre $P^\beta = P^{\alpha\gamma} = P$. $\alpha\beta\gamma$ ist also involutorisch und $(\alpha\beta\gamma)$ ein mit g inzidierender Punkt.

In jedem Falle existiert ein P mit $P\,\mathrm{I}\,g$, $P \mid \alpha, \beta$. (A5) liefert ein Q mit $PQ \neq QP$, $Q \mid \alpha, \beta$ also $Q\,\mathrm{I}\,g$ (5a). $P^Q = R$ ist nun ein von P und Q verschiedener Punkt, welcher ebenfalls mit g inzidiert.

Satz 7. *Es ist niemals $\alpha\beta\gamma = 1$, d.h. es ist niemals ein α gleich einem* g.

Beweis. Angenommen, es ist $\alpha\beta\gamma = 1$. Dann gilt $\alpha \mid \beta \mid \gamma$. Zu α, β gibt es nach (6) ein P mit $P \equiv (\alpha\beta\pi)$. Wegen $P = \gamma\pi$ gilt $\pi \mid \gamma$, damit $\alpha \mid \beta \mid \gamma \mid \pi$, also (1) $\alpha\beta\gamma = \pi = 1$ (Widerspruch).

Allein aus den Definitionen gewinnt man die folgenden Sätze 8 und 9.

Satz 8. *Aus $P\,\mathrm{I}\,g$, $g\,\mathrm{I}\,\varepsilon$ folgt $P\,\mathrm{I}\,\varepsilon$.*

Beweis. Nach Voraussetzung gibt es $\alpha, \beta, \gamma, \delta$ derart, daß $P \equiv (\alpha\beta\gamma)$, $g \equiv (\alpha\beta) \equiv (\varepsilon\delta)$ ist, und daß aus $\sigma \mid \alpha, \beta$ folgt $\sigma \mid \varepsilon, \delta$. Damit gilt $\gamma \mid \delta$ und $P\varepsilon = \alpha\beta\gamma\varepsilon = \varepsilon\delta\gamma\varepsilon = (\delta\gamma)^\varepsilon$ ist involutorisch.

Satz 9. *Aus $\alpha\,\mathrm{I}\,g$, $g \perp \varepsilon$ folgt $\alpha \perp \varepsilon$.*

Beweis. Nach Voraussetzung ist g darstellbar als $g \equiv (\alpha\beta)$ und als $g \equiv (\gamma\delta)$ mit $\gamma, \delta \mid \varepsilon$. $(\alpha\beta) \equiv (\gamma\delta)$ ergibt aber, daß auch $\alpha, \beta \mid \varepsilon$ gilt.

Die folgenden drei Sätze handeln von den grundlegenden Eigenschaften der Inzidenz von Punkten, Geraden und Ebenen.

Satz 10. *Zu P, Q gibt es ein g mit $g\,\mathrm{I}\,P, Q$.*
(*) *Aus $g, h\,\mathrm{I}\,P, Q$ folgt $g \equiv h$ oder $P \equiv Q$.*

Beweis. Nach (A1) gibt es ein α mit $\alpha \mid P, Q$; nach (4) ein β mit $\beta \mid P, Q, \alpha$. Nun sind (5a) P, Q darstellbar als $P \equiv (\alpha\beta\vartheta)$, $Q \equiv (\alpha\beta\pi)$, und $g \equiv (\alpha\beta)$ ist eine Gerade mit $g\,\mathrm{I}\,P, Q$.

(*) Es sei $P \neq Q$; $P, Q \mathrm{I} (\alpha\beta), (\gamma\delta)$. Dann gilt $P, Q \mid \alpha, \beta, \gamma, \delta$ (8); $\alpha\beta\gamma = \delta'$ (A3); $\delta' \mid \gamma, P, Q$ (2a) und schließlich $\delta' = \delta$ (4*). Es ist also $\alpha\beta = \gamma\delta$. Gilt nun $\tau \mid \alpha, \beta$, so folgt $(\alpha\beta\tau) \equiv T$ (7). Aus $P \neq Q$, $\alpha \neq \beta$; $P, Q \mid \alpha, \beta, \gamma, \delta$; $T \mid \alpha, \beta$ ergibt sich mit (A2) $T \mid \gamma, \delta$; $T \equiv (\gamma\delta\tau')$ (5a). Wegen $\alpha\beta = \gamma\delta$ ist aber $\tau' = \tau$ und damit $\tau \mid \gamma, \delta$. In gleicher Weise zeigt man umgekehrt, daß aus $\sigma \mid \gamma, \delta$ folgt $\sigma \mid \alpha, \beta$, d.h. aber, daß $(\alpha\beta) \equiv (\gamma\delta)$ gilt.

Satz 11. *a) Aus* $P \neq Q$; $P, Q \mathrm{I} \varepsilon, g$ *folgt* $\varepsilon \mathrm{I} g$.

b) Gilt $P \mathrm{I} \alpha, \beta$, *so gibt es ein* g *mit* $g \mathrm{I} P, \alpha, \beta$.

Beweis. a) Es sei $g \equiv (\alpha\beta)$. Dann gilt $P, Q \mid \alpha, \beta$ (8); $\varepsilon\alpha\beta = \delta$ (A3); $\delta \mid P, Q$ (2a); $P, Q \mathrm{I} (\varepsilon\delta)$ (5a); $(\alpha\beta) \equiv (\varepsilon\delta)$ (10*) und damit $\varepsilon \mathrm{I} g$.

b) Nach (A5) gibt es ein Q mit $P \neq Q$; $Q \mid \alpha, \beta$. Die Verbindungsgerade von P und Q (10) inzidiert (a) dann auch mit α und β.

Satz 12. *a) Zu* P, Q, R *gibt es ein* ε *mit* $\varepsilon \mathrm{I} P, Q, R$.

a) Aus* $P, Q, R \mathrm{I} \varepsilon, \delta$ *folgt* $\varepsilon = \delta$ *oder* P, Q, R *sind kollinear*.

b) Zu g, P *gibt es ein* ε *mit* $\varepsilon \mathrm{I} g, P$.

b) Aus* $g, P \mathrm{I} \varepsilon, \delta$ *folgt* $\varepsilon = \delta$ *oder* $g \mathrm{I} P$.

c) Gilt $g, h \mathrm{I} P$, *so gibt es ein* ε *mit* $g, h \mathrm{I} \varepsilon$.

c) Aus* $g, h \mathrm{I} \varepsilon, \delta$ *folgt* $\varepsilon = \delta$ *oder* $g \equiv h$.

Beweis. a) ist gleichbedeutend mit (A1).

a*) Zu P, Q gibt es ein g mit $g \mathrm{I} P, Q$ (10), welches als $g \equiv (\varepsilon\alpha)$ darstellbar ist (4). Ist $P \neq Q$, $\varepsilon \neq \delta$, so folgt $\alpha \mid R$ (A2) und damit $R \mathrm{I} g$ (5a). Gilt aber $P = Q$, so sind P, Q, R nach (10) kollinear.

b) R, Q seien verschiedene Punkte mit $R, Q \mathrm{I} g$ (6). Ein ε mit $\varepsilon \mathrm{I} P, Q, R$ (a) inzidiert nach (11a) auch mit g.

b*) ε, δ inzidieren mit zwei verschiedenen Punkten Q, R auf g (6) (8). Ist $\varepsilon \neq \delta$, so liegen P, Q, R auf einer Geraden (a*), welche nach (10*) mit g identisch ist.

c) Nach (6) gibt es zwei Punkte Q, R mit $P \neq Q$, $P \neq R$, $Q \mathrm{I} g$, $R \mathrm{I} h$. Ein ε mit $\varepsilon \mathrm{I} P, Q, R$ (a) inzidiert nach (11a) auch mit g und h.

c*) Gilt $g \neq h$, so haben g, h höchstens einen Punkt gemein (10*), es gibt also ein P mit $P \mathrm{I} h$, $P \mathrm{I} g$ (6). Nun folgt $\varepsilon, \delta \mathrm{I} g, P$ (8) also $\varepsilon = \delta$ (b*).

Eine Erweiterung von Satz 7 ist der

Satz 13. *Es ist niemals ein* P *gleich einem* $\mathfrak{g}$.

Beweis. Angenommen, es ist $P = \mathfrak{g} = \sigma\tau$ $(\sigma \mid \tau)$. Zu $P, g \equiv (\sigma\tau)$ gibt es nach (12b) ein ε mit $\varepsilon \mathrm{I} P, g$; d.h. $g \equiv (\varepsilon\gamma)$, $P \equiv (\alpha\beta\varepsilon)$ (3). Damit folgt $\alpha\beta = \gamma$, $\alpha\beta\gamma = 1$ im Widerspruch zu (7).

Der nächste Satz sagt aus, daß polare Geraden windschief sind:

Satz 14. *Aus* $\mathfrak{g} = \mathfrak{h}$; $g, h \mathrm{I} \varepsilon$ *oder* $g, h \mathrm{I} P$ *oder* $g, h \perp \varepsilon$ *folgt* $g \equiv h$.

Beweis. $g, h \mathrm{I} \varepsilon$ bedeutet: Es gibt Darstellungen $g \equiv (\varepsilon\alpha)$, $h \equiv (\varepsilon\beta)$. $\mathfrak{g} = \mathfrak{h}$ ergibt dann $\alpha = \beta$. Gilt $g, h \mathrm{I} P$, so gibt es (12c) ein ε mit $g, h \mathrm{I} \varepsilon$. Gilt $g, h \perp \varepsilon$, $g \equiv (\alpha\beta)$, $h \equiv (\gamma\delta)$, so ist $(\alpha\beta\varepsilon) \equiv (\gamma\delta\varepsilon)$ ein mit g und h inzidierender Punkt (7).

Wir zeigen nun, daß es stets möglich ist, von einem Punkt auf eine Ebene ein „*Lot*" und von einer Geraden auf eine Ebene eine „*Lotebene*" zu fällen.

Satz 15. *a) Zu P, ε gibt es ein l mit $l \mathrm{I} P$, $l \perp \varepsilon$.*
a) Aus $l, m \mathrm{I} P$; $l, m \perp \varepsilon$ folgt $l \parallel m$ oder $P = \varepsilon$.*
b) Zu g, ε gibt es ein λ mit $\lambda \mathrm{I} g$, $\lambda \perp \varepsilon$.
b) Aus $\lambda, \mu \mathrm{I} g$; $\lambda, \mu \perp \varepsilon$ folgt $\lambda = \mu$ oder $g \perp \varepsilon$.*

Beweis. a) 1. Fall: $P^\varepsilon \neq P$: Nach (10) gibt es ein l mit $l \mathrm{I} P$, P^ε. Wegen $P \not\mathrm{I} \varepsilon$ haben l und ε höchstens einen Punkt gemein (8) (11a), daher gibt es (A6) (A5) ein Q mit $Q \mathrm{I} \varepsilon$, $Q \not\mathrm{I} l$. Sei $\lambda \mathrm{I} P$, P^ε, Q (12a), so gilt $\lambda^\varepsilon \mathrm{I} P^\varepsilon$, P, Q und daher $\lambda^\varepsilon = \lambda$ (12a*). Mit $l^\varepsilon \mathrm{I} P^\varepsilon$, P folgt $l^\varepsilon = l$ (10*). Ist $l = (\lambda\mu)$ (11a), so gilt $\mu^\varepsilon = (\lambda l)^\varepsilon = \lambda^\varepsilon l^\varepsilon = \lambda l = \mu$, und aus $\lambda, \mu \mathrm{I} P$ (8), $\varepsilon \not\mathrm{I} P$ ergibt sich $\lambda, \mu \mid \varepsilon$ also $l \perp \varepsilon$.

2. Fall: $P \mid \varepsilon$: Es folgt $P = (\varepsilon\lambda\mu)$ (3), $(\lambda\mu) \mathrm{I} P$, $(\lambda\mu) \perp \varepsilon$.

3. Fall: $P = \varepsilon$: Ist $P = (\alpha\lambda\mu)$, so gilt $\lambda, \mu \mid \varepsilon$ also $(\lambda\mu) \perp \varepsilon$, $(\lambda\mu) \mathrm{I} P$.

a*) 1. Fall: $P^\varepsilon \neq P$: Es folgt $l = l^\varepsilon$, $m = m^\varepsilon \mathrm{I} P$, P^ε also $l = m$ (10*).

2. Fall: $P \mid \varepsilon$: Ist $l = (\lambda\mu)$, $m = (\lambda'\mu')$, so folgt $P = (\lambda\mu\varepsilon) = (\lambda'\mu'\varepsilon)$ (5b) also $\lambda\mu = \lambda'\mu'$ und damit (14) $l = m$.

b) Nach (6) gibt es ein P mit $P \mathrm{I} g$, nach (a) ein l mit $l \mathrm{I} P$, $l \perp \varepsilon$, nach (12c) ein λ mit $\lambda \mathrm{I} l, g$. Satz (9) ergibt aber $\lambda \perp \varepsilon$.

b*) Da $\mathfrak{g} = \varepsilon$ unmöglich ist (13), haben wir drei Fälle zu unterscheiden:

1. Fall: $\mathfrak{g}^\varepsilon \neq \mathfrak{g}$: Es folgt $\lambda = \lambda^\varepsilon$, $\mu = \mu^\varepsilon \mathrm{I} g^\varepsilon$ und damit $\lambda = \mu$ (12c*).

2. Fall: $g \mathrm{I} \varepsilon$: Es sei $g = (\varepsilon\alpha)$. Dann folgt $\varepsilon, \lambda \mathrm{I} g$, $(\varepsilon\lambda)$; $\varepsilon, \mu \mathrm{I} g$, $(\varepsilon\mu)$, damit (12c*) $g = (\varepsilon\lambda) = (\varepsilon\mu)$ also $\lambda = \mu$.

3. Fall: $\mathfrak{g} \mid \varepsilon$, $g \not\mathrm{I} \varepsilon$: Es sei $g = (\lambda\nu)$. Dann gilt $\nu\varepsilon = \lambda\lambda\nu\varepsilon = \lambda\varepsilon\lambda\nu = \varepsilon\nu$. Aus $\nu = \varepsilon$ folgte $g \mathrm{I} \varepsilon$, also ist $\nu \mid \varepsilon$ d.h. $g \perp \varepsilon$.

Ein später noch oft benötigter Hilfssatz ist der

Satz 16. *Aus $g, \alpha \perp \varepsilon$; $g, \alpha \mathrm{I} P$ folgt $g \mathrm{I} \alpha$ oder $P = \varepsilon$.*

Beweis. Nach (15b) gibt es ein λ mit $\lambda \mathrm{I} g$, $\lambda \perp \alpha$. Es folgt (8) $\lambda \mathrm{I} P$, $\lambda \perp \varepsilon$ (9), $(\alpha\lambda) \perp \varepsilon$, $(\alpha\lambda) \mathrm{I} P$ (5a) und damit $g = (\alpha\lambda)$ oder $P = \varepsilon$ (15a*).

Eine Ergänzung zu Satz 12 ist der

Satz 17. *Gilt $g, h \perp \varepsilon$, so gibt es ein ω mit $\omega \mathrm{I} g, h$; $\omega \perp \varepsilon$.*

Beweis. Ist $g = (\alpha\beta)$, so ist $(\alpha\beta\varepsilon)$ ein Punkt P (7). Zu P, h gibt es (12b) ein ω mit $\omega \mathrm{I} P, h$. Nach (9) folgt $\omega \perp \varepsilon$, und wegen $P \neq \varepsilon$ ergibt (16) $\omega \mathrm{I} g$.

Im folgenden Satz verstehen wir unter g, h jeweils beliebige, zu $\mathfrak{g}, \mathfrak{h}$ gehörige Geraden, d.h. Geraden, deren Spiegelungen $\mathfrak{g}, \mathfrak{h}$ sind.

Satz 18. *a) Aus $\varepsilon \mid \delta$ folgt $\varepsilon \perp \delta$.*
b) Aus $\varepsilon \mid P$ folgt $\varepsilon \mathrm{I} P$.
c) Aus $\varepsilon \mid \mathfrak{g}$ folgt $\varepsilon \mathrm{I} g$ oder $\varepsilon \perp g$.
d) Gilt $P \mid Q$, so gibt es ein π mit $P = \pi$, $Q \mathrm{I} \pi$.
e) Gilt $P \mid \mathfrak{g}$, so folgt $g \mathrm{I} P$ oder $g \mathrm{I} \pi$ mit $\pi = P$.
f) Gilt $\mathfrak{g} \mid \mathfrak{h}$; $g, h \mathrm{I} P$ oder $g, h \mathrm{I} \varepsilon$; so folgt $g \perp h$.

Beweis. a) und b) sind Definitionen. c) Es sei $\lambda \mathrm{I} g$, $\lambda \perp \varepsilon$, $g = (\lambda\mu)$ (15b). Es folgt $\mu\varepsilon = \varepsilon\mu$ also $\varepsilon = \mu$, $g \mathrm{I} \varepsilon$ oder $\varepsilon \mid \mu$, $g \perp \varepsilon$.

d) Ist $(\alpha\beta) \mathrm{I} P$, Q (10), $P = (\alpha\beta\vartheta)$, $Q = (\alpha\beta\pi)$ (5a), so folgt $\vartheta\pi = PQ$ ist involutorisch und damit $\alpha \mid \beta \mid \vartheta \mid \pi$ also (1) $\alpha\beta\vartheta\pi = 1$, $P = \pi$, $Q \mathrm{I} \pi$.

e) Es sei $\gamma \mathrm{I} P, g$ (12b), $P \equiv (\alpha\beta\gamma)$, $g \equiv (\pi\gamma)$ (3). Aus $\mathfrak{g} \mid P$ folgt: $\alpha\beta\pi$ ist involutorisch und damit $\pi\alpha\beta\gamma = \gamma\beta\alpha\pi$. Ist nun $\pi\alpha\beta\gamma = 1$, so gilt $\pi = P$, $g \mathrm{I} \pi$; ist dagegen $\pi\alpha\beta\gamma$ involutorisch, so ergibt sich $P \nmid \pi$ und damit $P \mathrm{I} g$ (5a).

f) Gilt $g, h \mathrm{I} P$, so gibt es auch ein ε mit $g, h \mathrm{I} \varepsilon$ (12c). Aus $g \equiv (\varepsilon\alpha)$, $h \equiv (\varepsilon\beta)$, $\mathfrak{g} \mid \mathfrak{h}$ folgt aber $\alpha \mid \beta$ und damit $g \perp h$.

Es sei bemerkt, daß aus $\mathfrak{g} \mid \mathfrak{h}$ allein nicht $g \perp h$ folgt. Es gibt im klassischen elliptischen Raum „involutorische Schiebungen", das sind involutorische Bewegungen, welche sich als Produkt zweier Geradenspiegelungen darstellen lassen, jedoch selbst keine Geradenspiegelungen sind. Diesem Phänomen liegt die Tatsache zugrunde, daß es in einer orthogonalen Gruppe O_4 Elemente geben kann, deren Quadrat das von der Identität verschiedene Element des Zentrums der O_4 ist.

Wir betrachten nun noch einmal polare Geraden:

Satz 19. *Sind g, h polar, so ist $g \mathrm{I} \varepsilon$ äquivalent zu $h \perp \varepsilon$.*
(*) *Aus $\alpha \neq \beta$; $\alpha, \beta \mathrm{I} g$; $\alpha, \beta \perp h$ folgt: g, h sind polar.*

Beweis. Aus $g \mathrm{I} \varepsilon$ folgt $\mathfrak{h} \mid \varepsilon$ also $h \mathrm{I} \varepsilon$ oder $h \perp \varepsilon$ (18c), aus $h \perp \varepsilon$ folgt $\mathfrak{g} \mid \varepsilon$ also $g \mathrm{I} \varepsilon$ oder $g \perp \varepsilon$. Wegen $g \not\equiv h$ ist aber $g, h \mathrm{I} \varepsilon$ und auch $g, h \perp \varepsilon$ ausgeschlossen (14) (17).

(*) Es sei $g \equiv (\alpha\varepsilon)$, $h \equiv (\gamma\delta)$. Nach (9) folgt $\gamma, \delta \mid \alpha, \beta$. Damit sind α, β Lotebenen durch g auf γ, δ und $\alpha \neq \beta$ ergibt (15b*) $g \perp \gamma, \delta$, d.h. $\alpha, \varepsilon \mid \gamma, \delta$ (9), $\alpha\varepsilon\gamma\delta = 1$ (1), $\mathfrak{g} = \mathfrak{h}$. Wegen $\alpha \mid \gamma, \delta$; $\alpha \nmid \alpha$ ist aber $g \not\equiv h$.

Wir wenden uns jetzt wieder dem Fällen von Loten und Lotebenen zu.

Satz 20. *Zu P, α, β gibt es ein λ mit $\lambda \mathrm{I} P$; $\lambda \perp \alpha, \beta$.*
(*) *Gilt $\lambda \neq \mu$; $\lambda, \mu \mathrm{I} P$; $\lambda, \mu \perp \alpha, \beta$, so gibt es ein l mit $l \mathrm{I} P$; $l \perp \alpha, \beta$.*

Beweis. Es sei $l, m \mathrm{I} P$; $l \perp \alpha$, $m \perp \beta$ (15a). Dann gibt es ein λ mit $\lambda \mathrm{I} l, m$ (12c), und es folgt $\lambda \mathrm{I} P$ (8) sowie $\lambda \perp \alpha, \beta$ (9).

(*) λ, μ, P inzidieren nach (11b) mit einem l, und (15b*) ergibt: Es ist $\lambda = \mu$ oder $l \perp \alpha, \beta$.

Satz 21. *a) Zu P, g gibt es ein λ mit $\lambda \mathrm{I} P$; $\lambda \perp g$.*
a) Aus $\lambda, \mu \mathrm{I} P$; $\lambda, \mu \perp g$ folgt $\lambda = \mu$ oder $P \mathrm{I} h$, h polar zu g.*
b) Zu P, g gibt es ein l mit $l \mathrm{I} P$; $l \perp g$.
b) Aus $l, m \mathrm{I} P$; $l, m \perp g$ folgt $l \equiv m$ oder $P \mathrm{I} g$ oder $P \mathrm{I} h$, h polar zu g.*

Beweis. a) Es sei $g \equiv (\alpha\beta)$. Nach (20) gibt es ein λ mit $\lambda \mathrm{I} P$; $\lambda \perp \alpha, \beta$ d.h. aber $\lambda \perp g$.

a*) Nach (11b) inzidieren λ, μ, P mit einer Geraden h. Satz (19*) ergibt nun $\lambda = \mu$ oder g, h polar.

b) Es sei $g \equiv (\alpha\beta)$; $\alpha \mathrm{I} P, g$ (12b); $\lambda \mathrm{I} P$, $\lambda \perp g$ (a). Dann gilt (9) $\alpha, \beta \perp \lambda$, und $l \equiv (\alpha\lambda)$ steht senkrecht auf g und inzidiert mit P (5a).

b*) Es sei $g \equiv (\alpha\beta) \equiv (\gamma\delta)$, $l \equiv (\alpha\lambda)$, $m \equiv (\gamma\mu)$ mit $\lambda \mid \beta$, $\mu \mid \delta$. Dann folgt $P \mathrm{I} \alpha, \gamma, \lambda, \mu$ (8). Inzidiert nun P nicht mit g, so ist $\alpha = \gamma$ (12b*), inzidiert P nicht mit einer Polaren von g, so ist $\lambda = \mu$ (a*).

Die Axiome 3 und 4 werden nun erweitert zu dem

Satz 22. *Satz von den drei Spiegelungen mit Umkehrung:*
a) Gilt $\alpha, \beta, \gamma \mathrm{I} g$, *so gibt es ein* δ *mit* $\delta = \alpha\beta\gamma$, $\delta \mathrm{I} g$.
a) Aus* $\alpha\beta\gamma = \delta$, $\alpha\beta \neq \beta\alpha$; $\alpha, \beta \mathrm{I} g$ *folgt* $\gamma \mathrm{I} g$.
b) Gilt $\alpha, \beta, \gamma \perp g$, *so gibt es ein* δ *mit* $\delta = \alpha\beta\gamma$, $\delta \perp g$.
b) Aus* $\alpha\beta\gamma = \delta$, $\alpha\beta \neq \beta\alpha$; $\alpha, \beta \perp g$ *folgt* $\gamma \perp g$.

Beweis. a) Nach (6) gibt es auf g zwei verschiedene Punkte P und Q. Es folgt: $\alpha, \beta, \gamma \,|\, P, Q$ (8), $\alpha\beta\gamma = \delta$ (A3), $\delta \,|\, P, Q$ (2a) also $\delta \mathrm{I} g$ (11a).

a*) Wir fällen eine Lotebene β' durch g auf γ. Mit (a) folgt $\alpha\beta\beta' = \alpha'$, $\alpha' \mathrm{I} g$. Die Gerade $(\beta'\gamma)$ heiße h. Wegen $\alpha\beta = \alpha'\beta'$ gilt $\alpha'\beta'\gamma = \delta$, $\beta'\gamma = \alpha'\delta = \mathfrak{h}$.

1. Fall: $(\alpha'\delta) \equiv h$. Dann ist $\alpha', \beta' \mathrm{I} g, h$, und aus $\alpha' \neq \beta'$ folgt (12c*) $g \equiv h$, $\gamma \mathrm{I} g$.

2. Fall: $(\alpha'\delta)$ polar zu h: Dann ist $\alpha' \perp h$ (19) also $\alpha' \,|\, \beta'$ (9), $\alpha \,|\, \beta$, $\alpha\beta = \beta\alpha$:

b) Es sei $g \equiv (\pi\vartheta)$. Dann folgt $\alpha, \beta, \gamma \,|\, \pi, \vartheta$ (9), und (A4) und (2b) ergeben $\alpha\beta\gamma = \delta$; $\delta \perp \pi, \vartheta$ also $\delta \perp g$.

b*) Nach (A6) gibt es zu γ ein P mit $P \mathrm{I} \gamma$, nach (21a) zu P, g ein γ' mit $\gamma' \mathrm{I} P$, $\gamma' \perp g$. Mit (b) ergibt sich $\alpha\beta\gamma' = \delta'$, $\delta' \perp g$. Eine Schnittgerade von γ und γ' (11b) heiße h.

1. Fall: $\gamma = \gamma'$. Wegen $\gamma' \perp g$ ist dann $\gamma \perp g$.

2. Fall: $\gamma \,|\, \gamma'$. Wegen $\gamma\delta = \gamma'\delta' = \beta\alpha$ sind dann $(\delta\delta') \equiv j$ und $(\gamma\gamma') \equiv h$ (12c*) polar oder identisch. Wäre j polar zu h, so folgte (19) $\delta \perp h$, $\delta \,|\, \gamma$ also auch $\alpha \,|\, \beta$, $\alpha\beta = \beta\alpha$. Es ist also $h \equiv j$ und damit $\delta' \mathrm{I} h$. Wegen $\gamma', \delta' \perp g$; $\gamma', \delta' \mathrm{I} h$; $\gamma'\delta' = \beta\alpha \neq 1$ ist nun h polar zu g (19*), und aus $\gamma \mathrm{I} h$ folgt $\gamma \perp g$ (19).

3. Fall: $\gamma\gamma' \neq \gamma'\gamma$. Dann folgt nach (a*), aus $\gamma\gamma'\delta' = \delta$; $\gamma, \gamma' \mathrm{I} h$, daß $\delta' \mathrm{I} h$ gilt, und damit ergibt sich wieder, daß h polar zu g (19*) und $\gamma \perp g$ (19) ist.

Aus Satz 22 gewinnt man leicht Aussagen über sog. gemischte Büschel, für die die nächsten zwei Sätze Beispiele sind.

Satz 23. *Gilt* $A, C \mathrm{I} g$; $\beta \perp g$, *so gibt es ein* δ *mit* $A\beta C = \delta$, $\delta \perp g$.
(*) *Aus* $AC \neq CA$; $A, C \mathrm{I} g$; $A\beta C = \delta$ *folgt* $\beta \perp g$.

Beweis. Es sei $g \equiv (\pi\vartheta)$, $A \equiv (\pi\vartheta\alpha)$, $C \equiv (\pi\vartheta\gamma)$. Wegen $\alpha, \beta, \gamma \perp g$ ergibt Satz (22b): $A\beta C = \mathfrak{g}\alpha\beta\gamma\mathfrak{g} = \alpha\beta\gamma = \delta$, $\delta \perp g$.

(*) Es sei wieder $g \equiv (\pi\vartheta)$, $A \equiv (\pi\vartheta\alpha)$, $C \equiv (\pi\vartheta\gamma)$. Dann gilt $\delta = A\beta C = \alpha\beta\gamma$. Aus $\gamma, \alpha \perp g$; $\alpha\gamma \neq \gamma\alpha$ und $\gamma\alpha\beta = \delta^{\gamma}$ folgt aber $\beta \perp g$ (22b*).

Satz 24. *a) Gilt* $a, b, c \perp g$, *so gibt es ein* d *mit* $g \perp d$ *und* $abc = \mathfrak{d}$.
b) Gilt $a, b, c \mathrm{I} \varepsilon$, P, *so gibt es ein* d *mit* $d \mathrm{I} \varepsilon$, P *und* $abc = \mathfrak{d}$.

Beweis. a) Nach der Definition der Orthogonalität von Geraden lassen sich a, b, c darstellen als $a \equiv (\alpha\alpha')$, $b \equiv (\beta\beta')$, $c \equiv (\gamma\gamma')$ mit $\alpha, \beta, \gamma \mathrm{I} g$; $\alpha', \beta', \gamma' \perp g$ also $\alpha, \beta, \gamma \perp \alpha', \beta', \gamma'$ (9). (22a, b) ergibt nun $abc = \alpha\alpha'\beta\beta'\gamma\gamma' = \alpha\beta\gamma\alpha'\beta'\gamma' = \delta\delta'$, $\delta \mathrm{I} g$, $\delta' \perp g$, $\delta\delta' = \mathfrak{d}$ (9), und aus $g \equiv (\varepsilon\delta)$ folgt $\varepsilon \,|\, \delta'$ (9) und damit $g \perp (\delta\delta') \equiv d$.

b) Nach (8) gilt $P \mathrm{I} \varepsilon$. Ist g das Lot in P auf ε (15a), und sind α, β, γ Verbindungsebenen (12c) von a, b, c mit g, so folgt $\alpha, \beta, \gamma \perp \varepsilon$ (9), $a \equiv (\alpha\varepsilon)$,

$b \equiv (\beta\varepsilon)$, $c \equiv (\gamma\varepsilon)$ (12c*) und (22a) ergibt $\mathfrak{a}bc = \alpha\varepsilon\beta\varepsilon\gamma\varepsilon = \alpha\beta\gamma\varepsilon = \delta\varepsilon$, $\delta \mathrm{I} g$, $\delta \perp \varepsilon$ (9), und aus $\delta \mathrm{I} P$ (8) folgt $P\mathrm{I}(\varepsilon\delta) \equiv d$ (5a).

Zum Schluß dieses Abschnittes beweisen wir eine räumliche Fassung des Lotensatzes (vgl. [*1*, § 3, Satz 13]) von HJELMSLEV.

Satz 25. Lotensatz:

Ist $\mathfrak{a}\alpha = A$, $\mathfrak{c}\gamma = C$ *und* $\mathfrak{a}\beta\mathfrak{c}$ *eine Ebenenspiegelung* δ, *so gilt:*

(25): *Steht* δ *auf einer Verbindungsgeraden von* A, C *senkrecht, so ist* $\alpha\beta\gamma$ *eine Ebenenspiegelung.*

(25*): *Ist* $AC \neq CA$ *und* $\alpha\beta\gamma$ *eine Ebenenspiegelung, so steht* δ *auf der Verbindungsgeraden von* A, C *senkrecht.*

Beweis. Es ist $\alpha\beta\gamma = \alpha\mathfrak{a}\mathfrak{a}\beta\mathfrak{c}\mathfrak{c}\gamma = A\,\delta C$. $\alpha\beta\gamma$ ist also dann und nur dann eine Ebenenspiegelung, wenn $A\,\delta C$ eine Erzeugende ist, und daher folgen beide Behauptungen aus (23).

§ 2. Darstellung der Bewegungen

In diesem Paragraphen stellen wir für die Elemente der durch das Axiomensystem in § 1 definierten Bewegungsgruppe $\mathfrak{G}$ Normalformen auf. Damit zeigen wir dann, daß $\mathfrak{G}$ kein echtes Zentrum besitzt, und daß $\mathfrak{G}$ stets einen Normalteiler vom Index 2 enthält.

Wir werden im folgenden Sätze aus § 1 in soweit zitieren, wie es für das Verständnis der Beweise notwendig erscheint. Existenz und Eindeutigkeit von Verbindungsebenen und -geraden, von Loten und Lotebenen sowie die Sätze 6, 8, 11 aus § 1 sollen jedoch im weiteren stillschweigend verwendet werden.

Wir beginnen mit einer wichtigen Folgerung aus dem Lotensatz:

Satz 26. *Zu* P, α, γ *gibt es ein* β *mit* $\beta \mathrm{I} P$, *so daß* $\alpha\beta\gamma \in \mathfrak{S}$.

Beweis. Inzidiert P mit einer Schnittgeraden von α und γ, bzw. mit einem gemeinsamen Lot auf α und γ, so folgt die Behauptung aus (22a) bzw. aus (21a)(22b). Anderenfalls fällen wir durch P Lote a, c auf α, γ mit den Fußpunkten $A = \mathfrak{a}\alpha$, $C = \mathfrak{c}\gamma$. Die Verbindungsgerade v von A und C bestimmt dann zusammen mit P eindeutig eine Ebene ω. Ist δ eine Ebene mit $\delta \mathrm{I} P$, $\delta \perp v$, und ist $a \equiv (\omega\alpha')$, $c \equiv (\omega\gamma')$, so sind α', γ' und auch δ (9) drei zu ω senkrechte Ebenen, welche mit dem Lot durch P auf ω inzidieren (16). Satz (22) ergibt nun $\mathfrak{a}\delta\mathfrak{c} = \omega\alpha'\delta\omega\gamma' = \alpha'\delta\gamma' = \beta$; $\beta \mathrm{I} P$, und eine Anwendung des Lotensatzes (25) liefert: $\alpha\beta\gamma$ ist eine Erzeugende.

Mit Satz 26 läßt sich ein Produkt der Form $\alpha\beta$ stets durch ein Produkt $\alpha'\beta'$ ersetzen derart, daß α' (bzw. β') mit einem beliebig vorgegebenen P inzidiert. Man suche nämlich zu α, β ein γ mit $\gamma \mathrm{I} P$ auf, so daß $\alpha\gamma\beta$ eine Erzeugende δ ist. Es folgt dann: $\alpha\beta = \gamma\delta^{\beta} = \delta^{\alpha}\gamma$.

Satz 27. *Gilt* $\alpha, \beta, \gamma, \delta \mathrm{I} P$, *so gibt es* ξ, η *mit* $\alpha\beta\gamma\delta = \xi\eta$; $\xi, \eta \mathrm{I} P$.

Beweis. α, β und γ, δ schneiden sich in je einer Geraden durch P. Ist nun ω eine Verbindungsebene dieser beiden Geraden, so ergibt (22a): $\alpha\beta\gamma\delta = \alpha\beta\omega\omega\gamma\delta = \xi\eta$; $\xi, \eta \mathrm{I} P$.

Die Sätze 26 und 27 ermöglichen folgende Reduktion:

Satz 28. *Ein Produkt von 5 Erzeugenden läßt sich auf ein Produkt von 3 Erzeugenden reduzieren. Jedes Element von* $\mathfrak{G}$ *ist als Produkt von höchstens 4 Erzeugenden darstellbar.*

Beweis. Es sei $\Phi = \alpha\beta\gamma\delta\varepsilon$. Wir wählen ein P mit $P\,I\,\alpha$ aus und ersetzen $\beta\gamma$ durch $\beta'\gamma'$ mit $\beta' I P$. Sodann ersetzen wir $\gamma'\delta$ durch $\gamma''\delta'$ mit $\gamma'' I P$ und endlich $\delta'\varepsilon$ durch $\delta''\varepsilon'$ mit $\delta'' I P$ (26). Nun ist $\Phi = \alpha\beta'\gamma''\delta''\varepsilon'$, und mit $\alpha, \beta', \gamma'', \delta'' I P$ folgt nach (27): $\alpha\beta'\gamma''\delta'' = \xi\eta$ also $\Phi = \xi\eta\varepsilon'$.

Eine Bewegung heiße „*gerade*", wenn sie sich als Produkt einer geraden, „*ungerade*", wenn sie sich als Produkt einer ungeraden Anzahl von Erzeugenden darstellen läßt. Satz 28 lehrt, daß sich jede gerade Bewegung als Produkt von 4, jede ungerade Bewegung als Produkt von 3 Erzeugenden schreiben läßt, denn ist die Bewegung als Produkt von weniger als 3 Erzeugenden gegeben, so kann man sie mit einem $\varepsilon\varepsilon = 1$ multiplizieren. Eine weitere Normalform liefert nun der

Satz 29. *a) Jede ungerade Bewegung ist in der Form* $\mathfrak{g}\varepsilon$ *darstellbar.*

b) Jede gerade Bewegung ist in der Form $\mathfrak{g}\mathfrak{h}$ *darstellbar.*

Beweis. a) Es sei $\Phi = \alpha\beta\gamma$. Wir wählen ein P mit $P I \gamma$ und ersetzen $\alpha\beta$ durch $\alpha'\beta'$ mit $\beta' I P$ (26). γ und β' schneiden sich dann in einer Geraden h. Ist nun λ eine Lotebene durch h senkrecht zu α', so folgt $\lambda\beta'\gamma = \varepsilon$ (22a) und damit $\Phi = \alpha'\lambda\lambda\beta'\gamma = \mathfrak{g}\varepsilon$.

b) Es sei $\Phi = \alpha\beta\gamma\delta$. Zunächst gilt nach (a): $\Phi = (\alpha\beta\gamma)\delta = (\mathfrak{a}\varepsilon)\delta$. Auf einem zu $\mathfrak{a}$ gehörigen a liege der Punkt O. Dann läßt sich Φ nach (26) auch darstellen als $\Phi = \mathfrak{a}\varepsilon'\delta'$ mit $\varepsilon' I O$. Sei nun λ eine Ebene mit $\lambda I O$; $\lambda \perp \varepsilon', \delta'$ (20). Dann folgt: $\Phi = \mathfrak{a}\varepsilon'\lambda\lambda\delta' = \mathfrak{a}\mathfrak{b}\mathfrak{c}$, und aus $\varepsilon', \lambda I O$ ergibt sich $(\varepsilon'\lambda) \equiv b\,I O$. Damit ist Φ ein Produkt von 3 Geradenspiegelungen, wobei sich zwei zugehörige Geraden schneiden.

a, b und O liegen in einer Ebene ω. Ist l ein Lot durch O senkrecht zu c; λ eine Lotebene durch l senkrecht zu σ, $l \equiv (\lambda\mu)$, $O \equiv (\lambda\mu\pi)$, so ist $m \equiv (\mu\pi)$ eine Gerade durch O mit $m I \omega$ (16) und $m \perp l$. Mit (24b) folgt $\mathfrak{a}\mathfrak{b}\mathfrak{m} = \mathfrak{n}$, $n I O$, ω (für ein zu $\mathfrak{n}$ gehöriges n) und damit $\Phi = (\mathfrak{a}\mathfrak{b}\mathfrak{m})(\mathfrak{m}\mathfrak{l})(\mathfrak{l}\mathfrak{c}) = \mathfrak{n}\mathfrak{o}\mathfrak{p}$. Hierbei gehören zu $\mathfrak{o}$, $\mathfrak{p}$ Geraden o, p mit o, $p \perp l$.

Für ein auf einer Verbindungsebene τ von n, l in O errichtetes Lot q gilt $q \perp n, l$ (Beweis etwa für $q \perp n$: Sei $\xi I q, n$; $q \equiv (\xi\eta)$, so folgt $\xi, \eta \mid \tau$ aus $q \perp \tau$ (9). $n I \xi, \tau$; $\xi \perp \tau$ ergibt aber $n \equiv (\xi\tau)$). Damit gilt (24a) $\mathfrak{q}\mathfrak{o}\mathfrak{p} = \mathfrak{h}$, und wir erhalten schließlich $\Phi = (\mathfrak{n}\mathfrak{q})(\mathfrak{q}\mathfrak{o}\mathfrak{p}) = \mathfrak{g}\mathfrak{h}$.

Wir zeigen nun, daß gerade und ungerade Bewegungen stets verschieden sind und damit den

Satz 30. *Die Gruppe* $\mathfrak{G}$ *enthält als Normalteiler vom Index 2 die Gruppe der geraden Bewegungen aus* $\mathfrak{G}$.

Beweis. Wäre eine gerade Bewegung einer ungeraden gleich, so bestünde eine Gleichung der Form $\mathfrak{j}\varepsilon = \mathfrak{g}\mathfrak{h}$ (29), und es folgte $\varepsilon = \mathfrak{j}\mathfrak{g}\mathfrak{h} = \mathfrak{a}\mathfrak{b}$ (29). Wegen $\varepsilon \mid \mathfrak{a}$ ergäbe sich für ein zu $\mathfrak{a}$ gehöriges a: $\varepsilon I a$ oder $\varepsilon \perp a$ (18c). Im ersten Fall

wäre $\alpha\varepsilon$ ein α, im zweiten ein A. Hieraus folgte jedoch $\alpha = \mathfrak{b}$ bzw. $A = \mathfrak{b}$ im Widerspruch zu (7) bzw. zu (13).

Da also die geraden Bewegungen eine echte Untergruppe bilden, und da das Produkt zweier ungerader Bewegungen gerade ist, so ist die Gruppe der geraden Bewegungen Untergruppe vom Index 2 in $\mathfrak{G}$ also Normalteiler.

Wir zeigen jetzt, daß die Bewegungen eines metrischen Raumes eine Gruppe bilden, welche isomorph ist zu der dem metrischen Raum zugrunde liegenden axiomatisch gegebenen Gruppe $\mathfrak{G}$:

Satz 31. *Das Zentrum von $\mathfrak{G}$ besteht nur aus dem Einselement; die Gruppe der inneren Automorphismen von $\mathfrak{G}$ ist isomorph zu $\mathfrak{G}$.*

Beweis. a) Das Zentrum $\mathfrak{Z}$ enthält kein ε: Gilt $P \mid \varepsilon$ (A6), $P \equiv (\alpha\beta\varepsilon)$, so gibt es ein Q mit $Q \mid \alpha, \beta$; $QP \neq PQ$ (A5), und es folgt $Q\varepsilon \neq \varepsilon Q$.

b) $\mathfrak{Z}$ enthält kein P: Ist $P \equiv (\alpha\beta\gamma)$, so gibt es ein Q mit $Q \mid \alpha$; $QP \neq PQ$.

c) $\mathfrak{Z}$ enthält kein $\mathfrak{g}$: Ist $\mathfrak{g} = \alpha\beta$, $g \equiv (\alpha\beta)$, $P \equiv (\alpha\beta\gamma)$ (6), $Q \mid \beta, \gamma$; $QP \neq PQ$, so folgt $Q\alpha \neq \alpha Q$ und damit $Q\mathfrak{g} \neq \mathfrak{g}Q$.

d) $\mathfrak{Z}$ enthält kein $\mathfrak{g}\varepsilon$: Gilt $\mathfrak{g}\varepsilon \neq \varepsilon\mathfrak{g}$, so ist $(\mathfrak{g}\varepsilon)\varepsilon \neq \varepsilon(\mathfrak{g}\varepsilon)$; gilt dagegen $\mathfrak{g}\varepsilon = \varepsilon\mathfrak{g}$, so ist $\varepsilon \mid \mathfrak{g}$ (7) also (18c) $\mathfrak{g}\varepsilon = \delta$ oder $\mathfrak{g}\varepsilon = P$.

e) $\mathfrak{Z}$ enthält kein $\mathfrak{g}\mathfrak{h}$: Es sei $\mathfrak{g} = \alpha\beta$, $g \equiv (\alpha\beta)$ und h eine zu $\mathfrak{h}$ gehörige Gerade. Gilt $\mathfrak{g}\mathfrak{h} \neq \mathfrak{h}\mathfrak{g}$, so ist $(\mathfrak{g}\mathfrak{h})\mathfrak{g} \neq \mathfrak{g}(\mathfrak{g}\mathfrak{h})$; gilt dagegen $\mathfrak{g}\mathfrak{h} = \mathfrak{h}\mathfrak{g}$, so ist entweder α nicht mit $\mathfrak{g}\mathfrak{h}$ vertauschbar oder es ist $\mathfrak{h} \mid \alpha, \beta$. Aus $h \mathrm{I} \alpha, \beta$ folgt $g \equiv h$, $\mathfrak{g}\mathfrak{h} = 1$; aus $h \perp \alpha, \beta$ ergibt sich ebenfalls $\mathfrak{g}\mathfrak{h} = 1$, da dann h polar zu g ist (19*). Gilt aber $h \mathrm{I} \alpha$, $h \perp \beta$ (bzw. $h \perp \alpha$, $h \mathrm{I} \beta$), so folgt $\mathfrak{g} \mid \mathfrak{h}$ also (18f) $g \perp h$, und $\mathfrak{g}\mathfrak{h}$ ist eine Geradenspiegelung.

§ 3. Das Axiom vom Polartetraeder

In diesem Paragraphen nehmen wir eine Gabelung unseres Aufbaus vor. Wie schon mehrfach bemerkt, ist es durch das Axiomensystem nicht ausgeschlossen, daß zu verschiedenen Elementen des metrischen Raumes gleiche Spiegelungen gehören. Dieser Fall tritt genau dann ein, wenn ein Polartetraeder existiert, d.h. wenn es vier paarweise aufeinander senkrecht stehende Ebenen gibt. Wir unterscheiden daher zwischen Bewegungsgruppen, in welchen es vier Erzeugende gibt, deren zugehörige Ebenen ein solches Polartetraeder bilden (Axiom P) und Bewegungsgruppen, in denen das nicht der Fall ist (Axiom $\sim$P). Die Bewegungsgruppe des klassischen elliptischen Raumes gehört zum ersten, die Bewegungsgruppen des klassischen euklidischen und hyperbolischen Raumes zum zweiten Typ.

Axiom P. *Es gibt $\alpha, \beta, \gamma, \delta$ mit $\alpha \mid \beta \mid \gamma \mid \delta$.*

Wir vermerken im folgenden hinter der Satznummer das jeweils vorausgesetzte Zusatzaxiom.

Satz 32 (AP). *Zu jedem Element des metrischen Raumes gibt es ein polares Element. Der Übergang zu polaren Elementen ist eine umkehrbar eindeutige, inzidenztreue Abbildung der Mengen der Ebenen, Geraden und Punkte auf die Mengen der Punkte, Geraden und Ebenen.*

Beweis. Die Erzeugenden β, γ, δ aus (AP) stellen ein $A \equiv (\beta\gamma\delta)$ dar (7) und es gilt $A = \alpha$ (1).

a) Es sei P ein beliebiger Punkt. Ist dann $l \equiv (\lambda\mu)$ eine Verbindungsgerade von A, P, so gilt $\lambda, \mu \mid \alpha$ also $l \perp \alpha$ und damit (23) $P = A\alpha P = \pi$.

b) Es sei ε eine beliebige Ebene. Ist dann $l \equiv (\lambda\mu)$ ein Lot durch A $(A \equiv (\lambda\mu\sigma))$ senkrecht auf ε, so folgt $\lambda, \mu \mid \alpha$, und (A 4) ergibt $\varepsilon = A\alpha\varepsilon = (\lambda\mu)(\sigma\alpha\varepsilon) = \lambda\mu\tau$; $\tau \mid \lambda, \mu$ (2) also $\varepsilon = E$, $E \equiv (\lambda\mu\tau)$ (7).

c) Es sei g eine beliebige Gerade, $g \equiv (\alpha\beta)$. Die Verbindungsgerade h der Pole A, B von α, β (b) steht auf α, β senkrecht und ist nach (19*) polar zu g. Wir haben nun noch zu zeigen, daß es zu einer Geraden nicht mehr als eine Polare geben kann (Die entsprechende Behauptung für Punkte und Ebenen ist trivial): Ist $g \equiv (\alpha\beta)$ polar zu h und h', so folgt $h, h' \perp \alpha$ (19). Damit sind h, h' komplanar (17) und aus $\mathfrak{g} = \mathfrak{h} = \mathfrak{h}'$ ergibt sich (14) $h \equiv h'$.

Es bleibt noch die Inzidenztreue der Abbildung auf die polaren Elemente zu zeigen: Aus $P \mathrm{I} \varepsilon$ folgt für $\pi = P$, $E = \varepsilon$ sofort $\pi \mid E$, d. h. $\pi \mathrm{I} E$. Ist h zu g polar und gilt $\varepsilon \mathrm{I} g$, $E = \varepsilon$, so folgt $\mathfrak{h} \mid E$, und da h nicht mit ε inzidiert (14), ergibt (18e) $h \mathrm{I} E$. Ist h zu g polar und gilt $P \mathrm{I} g$, $\pi = P$, so gilt $g \perp \pi$, und aus $\mathfrak{h} \mid \pi$ folgt, da h nicht auf π senkrecht steht (17) (14), daß $h \mathrm{I} \pi$ gilt.

Satz 33 (AP). *Der zu einer Bewegungsgruppe* $\mathfrak{B}$, *welche dem Axiomensystem* (§ 1) *und dem Axiom P genügt, gehörende metrische Raum erfüllt die projektiven Inzidenzaxiome.*

Beweis. Das Axiomensystem von A. Winternitz [9] für den dreidimensionalen projektiven Raum lautet:

Axiom a). Liegen zwei verschiedene Punkte A und B in zwei verschiedenen Ebenen α und β, der Punkt C gleichfalls in α und β, während die Ebene γ durch A und B hindurchgeht, so liegt C in γ.

Axiom b). Zu irgend drei Punkten gibt es stets eine Ebene, welche durch sie hindurchgeht.

Axiom b).* Zu irgend drei Ebenen gibt es stets einen Punkt, welcher in ihnen liegt.

Axiom c). Man kann fünf Punkte P_1, P_2, P_3, P_4, P_5 und fünf Ebenen $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5$ angeben, so daß P_i dann und nur dann außerhalb ε_k liegt, wenn $i - k \equiv \pm 2 \pmod 5$ ist.

Axiom a) ist genau unser Axiom 2, Axiom b) unser Axiom 1 in der Sprechweise des metrischen Raumes. Axiom b*) ist dual zu Axiom b) und kann daher mit Satz (32) aus (A 1) gefolgert werden. Eine Grundfigur nach Axiom c) läßt sich in folgender Weise herstellen: In einem Polartetraeder $\alpha, \beta, \gamma, \delta$ mit der Ecke $A \equiv (\beta\gamma\delta)$ suche man auf den Kanten $(\beta\gamma)$, $(\beta\delta)$, $(\gamma\delta)$ je einen Punkt auf, dessen Spiegelung nicht mit A kommutiert (A 5). $\alpha, \beta, \gamma, \delta$ zusammen mit der Verbindungsebene der drei erhaltenen Punkte ergeben fünf Ebenen, aus welchen sich durch Hinzufügen entsprechend ausgewählter Schnittpunkte die in Axiom c) geforderte Figur herstellen läßt.

Wir untersuchen nun die Alternative zu Axiom P:

Axiom ~P. *Aus* $\alpha \mid \beta, \gamma, \delta$; $\beta \mid \gamma, \delta$ *folgt* $\gamma \nmid \delta$.

Zwei zueinander polare Elemente eines metrischen Raumes werden stets durch insgesamt vier Ebenen dargestellt, die ein Polartetraeder bilden (Für ein Paar Pol-Polarebene folgt dies in trivialer Weise aus den Definitionen, für ein Paar polarer Geraden aus Satz (19)). Unter Voraussetzung von Axiom $\sim$P gibt es daher überhaupt keine polaren Elemente. Eine Gerade ist dann durch die zugehörige Geradenspiegelung eindeutig bestimmt. Wir werden im folgenden in allen Sätzen und Beweisen, in welchen Axiom $\sim$P vorausgesetzt ist, für Geraden und Geradenspiegelungen gemeinsam kleine lateinische Buchstaben verwenden.

Zunächst durchmustern wir die Sätze aus § 1 und § 2 und geben die Modifikationen an, welche sich unter Voraussetzung von (A$\sim$P) ergeben:

Satz 15′ (A$\sim$P). a*) *Das Lot von P auf ε ist eindeutig bestimmt.*

Satz 16′ (A$\sim$P). *Aus $g, \alpha \perp \varepsilon$; $g, \alpha \mathrm{I} P$ folgt stets $g \mathrm{I} \alpha$.*

Satz 18′ (A$\sim$P). d) *Es ist niemals $P \mid Q$.*

e) *Aus $P \mid g$ folgt $P \mathrm{I} g$.*

f) *Aus $g \mid h$ folgt $g \perp h$.*

Beweis zu 18′f. Es sei $P \mathrm{I} g$, $P \not\mathrm{I} h$, $Q \not\mathrm{I} g$, $Q \mathrm{I} h$ (6); $\varepsilon \mathrm{I} P, Q, P^h$. Aus $P^h \mid g$ folgt (18′e) $P^h \mathrm{I} g$ also $g \mathrm{I} \varepsilon$, $g \equiv (\varepsilon \delta)$, und $g \mid h$; $\varepsilon^h \mathrm{I} P^h, Q, P$ ergibt $\varepsilon^h = \varepsilon$, $\delta^h = \delta$. Wäre $h \perp \varepsilon, \delta$, so wäre h zu g polar (19*). $h \mathrm{I} \varepsilon$ oder $h \mathrm{I} \delta$ ergibt aber $h \perp g$ (18f).

Satz 19′ (A$\sim$P). *Gilt $\alpha, \beta \perp h$, so schneiden sich α und β nicht (unter Voraussetzung von Axiom $\sim$P gelangt man niemals zu einem metrischen Raum, der den projektiven Inzidenzaxiomen genügt).*

Satz 21′ (A$\sim$P). a*) *Die Lotebene von P auf g ist eindeutig bestimmt.*

b*) *Gilt $P \not\mathrm{I} g$, so ist das Lot von P auf g eindeutig bestimmt.*

Satz 22′ (A$\sim$P). *In* (22a*) *und* (22b*) *kann die Voraussetzung $\alpha\beta \neq \beta\alpha$ durch die schwächere Voraussetzung $\alpha \neq \beta$ ersetzt werden.*

Beweis. Der 2. Fall im Beweis von (22a*) sowie der 2. und der 3. Fall im Beweis von (22b*) führen auf die Existenz polarer Geraden.

Satz 23′ (A$\sim$P). (*) *Die Voraussetzung $AC \neq CA$ ist durch $A \neq C$ zu ersetzen.*

Satz 25′ (A$\sim$P). (*) *Die Voraussetzung $AC \neq CA$ ist durch $A \neq C$ zu ersetzen.*

Zu Satz 26 fügen wir eine Eindeutigkeitsaussage hinzu:

Satz 34 (A$\sim$P). *Gilt $\alpha \neq \gamma$, $\alpha\beta\gamma \in \mathfrak{S}$, $\alpha\beta'\gamma \in \mathfrak{S}$, und ist P ein Punkt, der mit β und β' aber nicht sowohl mit α als auch mit γ inzidiert, so folgt $\beta = \beta'$.*

Beweis. Es sei $\alpha\beta\gamma = \delta$, $\alpha\beta'\gamma = \delta'$ also $\beta\beta' = \alpha\delta\gamma\gamma\delta'\alpha = (\delta\delta')^\alpha = \delta^\alpha \delta'^\alpha$. Angenommen, es ist $\beta \neq \beta'$. Dann ergibt (22′): $\delta^\alpha \mathrm{I} P$. Wegen $\beta\delta^\alpha = \beta\alpha\alpha\beta\gamma\alpha = \gamma\alpha$ gilt $\beta \neq \delta^\alpha$, und eine zweite Anwendung von (22′) liefert $\alpha, \gamma \mathrm{I} P$.

Abschließend sei eine Bemerkung über die involutorischen Elemente der axiomatisch gegebenen Bewegungsgruppen angefügt:

Satz 35 (AP). *Ein involutorisches Element aus $\mathfrak{G}$ ist entweder eine Punkt-Ebenen-Spiegelung oder eine Geradenspiegelung oder eine involutorische Schiebung.*

(A∼P) *Ein involutorisches Element aus* $\mathfrak{G}$ *ist entweder eine Ebenenspiegelung oder eine Punktspiegelung oder eine Geradenspiegelung.*

Beweis. Nach (29) ist jedes Element entweder als ein $\mathfrak{g}\varepsilon$ oder als ein $\mathfrak{g}\mathfrak{h}$ darstellbar. Produkte $\mathfrak{g}\mathfrak{h}$ mit $\mathfrak{g}\,|\,\mathfrak{h}$ haben wir involutorische Schiebungen genannt, falls sie keine Geradenspiegelungen sind. Alles weitere ergibt sich unmittelbar aus Satz (18), (18′) und den Definitionen.

§ 4. Bewegungsgruppen metrischer Ebenen

Im folgenden Abschnitt wird ein Zusammenhang zu den von F. BACHMANN [*1*] betrachteten Bewegungsgruppen metrischer Ebenen hergestellt. Es zeigt sich, daß in den durch unser Axiomensystem bestimmten Gruppen stets Bewegungsgruppen metrischer Ebenen als Untergruppen (oder als Faktorgruppen von Untergruppen) enthalten sind. Insbesondere gilt, daß jede Ebene des metrischen Raumes eine metrische Ebene im Sinne von F. BACHMANN ist.

Satz 36. *Die Ebenenspiegelungen, welche zu Ebenen gehören, die mit einem festen Punkt O inzidieren, erzeugen eine Untergruppe* $\mathfrak{P}$ *von* $\mathfrak{G}$, *deren Faktorgruppe nach ihrem Zentrum* $\{O\}$ *isomorph zu einer ebenen elliptischen Bewegungsgruppe (im Sinne von* F. BACHMANN*) ist.*

Beweis. Unter $\alpha, \beta, \gamma, \delta$ verstehen wir in diesem Beweis Ebenen mit $\alpha, \beta, \gamma, \delta\, I O$. Nach (27) (29) sind die Elemente von $\mathfrak{P}$ darstellbar als $\alpha, \alpha\beta$ oder $\alpha\beta\gamma = \mathfrak{g}\varepsilon$. Damit liegen in $\mathfrak{P}$ außer O genau die folgenden involutorischen Elemente: Die Erzeugenden α und die Geradenspiegelungen $\mathfrak{g}$ mit $\mathfrak{g}\,|\,O$. Zu letzteren gehört jeweils genau eine Gerade g mit $g\,I\,O$ (polare Geraden sind punktfremd (14)). Die involutorischen Elemente von $\mathfrak{P}/\{O\}$ sind daher Restklassen $[\alpha; O\alpha]$ mit $\alpha \perp g$ für das zu $\mathfrak{g} = O\alpha$ gehörige g durch O. Nach [*1*, § 16] ist nun zu zeigen:

G) Kein $[\alpha; O\alpha]$ ist mit allen involutorischen Elementen von $\mathfrak{P}/\{O\}$ vertauschbar.

T) Ist $\alpha \neq \beta$, und sind $\alpha\beta\gamma$, $\alpha\beta\delta$ involutorisch, so ist $\alpha\gamma\delta$ involutorisch.

V) Zu α, β gibt es stets ein γ mit $\alpha, \beta \,|\, \gamma$.

Wir haben T) und V) nur für Ebenenspiegelungen ausgesprochen, der Übergang zu den Restklassen ist trivial.

Beweis von G). Ist P ein Punkt auf α mit $PO \neq OP$ (A5), l das Lot in P auf α, R ein Punkt mit $R\,I\,l$; $RP \neq PR$ und h eine Gerade mit $h\,I\,O$, R, so ist das zugehörige $\mathfrak{h}$ (und auch $\mathfrak{h}O = \beta$) nicht mit α vertauschbar (15a*).

Beweis von T). Gilt $\alpha\beta \neq \beta\alpha$, so inzidieren $\alpha, \beta, \gamma, \delta$ mit einer Geraden (11b) (22a*), und $\alpha\gamma\delta$ ist eine Erzeugende (22a). Ist aber $\alpha \,|\, \beta$ und inzidiert γ (bzw. δ) nicht mit der Geraden $(\alpha\beta)$, so steht γ (bzw. δ) auf $(\alpha\beta)$ senkrecht (18c) und $\alpha\beta\gamma$ (bzw. $\alpha\beta\delta$) ist gleich O.

Beweis von V). V) ergibt sich aus (20).

Satz 37 (A∼P). *Unter Voraussetzung von Axiom* ∼P *erzeugen die Ebenenspiegelungen, welche zu Ebenen gehören, die auf einer festen Ebene* ω *senkrecht*

stehen, eine Untergruppe $\mathfrak{E}$ von $\mathfrak{G}$, welche zu einer (nichtelliptischen) Bewegungsgruppe einer metrischen Ebene isomorph ist.

Beweis. Unter $\alpha, \beta, \gamma, \delta, \varepsilon; g, h, j$ verstehen wir in diesem Beweis Ebenen bzw. Geraden, welche auf ω senkrecht stehen. Mit α, β ist auch α^β senkrecht zu ω ($\alpha^\beta = \beta\alpha\beta$ kommutiert mit ε, wäre $\alpha^\beta = \varepsilon$, so folgte $\varepsilon^\beta = \alpha \neq \varepsilon$). Die Ebenenspiegelungen α mit $\alpha \mid \omega$ bilden also ein invariantes Erzeugendensystem von $\mathfrak{E}$. Den $P, Q, \ldots$ im Axiomensystem aus [*1*, § 3] entsprechen bei uns die Geraden (bzw. Geradenspiegelungen) $g, h, \ldots$. Damit erhalten die Bachmannschen Axiome folgende Gestalt:

1) Zu g, h gibt es stets ein α mit $g, h \mid \alpha$.

2) Aus $g, h \mid \alpha, \beta$ folgt $g = h$ oder $\alpha = \beta$.

3) Gilt $\alpha, \beta, \gamma \mid g$, so gibt es ein δ mit $\alpha\beta\gamma = \delta$.

4) Gilt $\alpha, \beta, \gamma \mid \varepsilon$, so gibt es ein δ mit $\alpha\beta\gamma = \delta$.

D) Es gibt α, β, γ derart, daß $\alpha \mid \beta$ und $\alpha \nmid \gamma$, $\beta \nmid \gamma$, $\gamma \nmid \alpha\beta$ gilt.

$\sim$P) (Axiom $\sim$P der Ebene). Es ist stets $\alpha\beta\gamma \neq 1$.

Da alle vorkommenden Geraden und Ebenen auf ω senkrecht stehen sollen, ist das $\mid$-Symbol zwischen einem α und einem g im metrischen Raum als Inzidenzzeichen zu deuten (18c) (19'). Damit folgt 1) aus (17) und 2) aus (12c*). Mit (22) und (A4) ergibt sich in 3) und 4) die Existenz eines δ mit $\alpha\beta\gamma = \delta$. Da ω mit α, β, γ kommutiert, gilt $\omega\delta = \delta\omega$. Wäre $\delta = \omega$, so folgte $\alpha \mid \beta \mid \gamma \mid \delta$ im Widerspruch zu (A$\sim$P), also ist $\delta \mid \omega$. Zum Beweis von D) betrachte man ein P auf ω (A6), $P \equiv (\omega\alpha\beta)$. Auf $(\omega\alpha)$ und $(\omega\beta)$ suche man je einen nicht mit P vertauschbaren Punkt auf (A5). Die Verbindungsebenen der Lote in den drei erhaltenen Punkten auf ω erfüllen dann die Bedingungen von D). Das Axiom $\sim$P der Ebene folgt schließlich aus Satz (7).

Unter Voraussetzung von Axiom P erzeugen die zu den auf einem ω senkrecht stehenden Ebenen gehörigen Spiegelungen eine Gruppe $\mathfrak{E}$, die die Grundebene ω mit enthält. Diese Gruppe ist aber identisch mit der nach Satz (36) durch den Pol von ω bestimmten Gruppe $\mathfrak{P}$. Die Faktorgruppe $\mathfrak{P}/\{\omega\}$ ist eine elliptische ebene Bewegungsgruppe. In beiden Fällen ($\sim$P und P) sind daher die Gruppen der inneren Automorphismen der Gruppen $\mathfrak{E}$ Bewegungsgruppen metrischer Ebenen, d.h. jede Ebene eines metrischen Raumes ist eine metrische Ebene im Sinne von F. BACHMANN.

Nach einem Satz aus [*1*] gibt es keine endlichen elliptischen Bewegungsgruppen der Ebene. Da die durch unser Axiomensystem gekennzeichneten Gruppen nach (36) stets ebene elliptische Bewegungsgruppen enthalten, kann es keine endlichen metrischen Räume geben.

§ 5. Büschel und Bündel

1. Ebenenbüschel. In § 3 wurde gezeigt, daß unter Voraussetzung von Axiom P für die Bewegungsgruppe $\mathfrak{G}$ der zugehörige metrische Raum ein projektiver Raum ist. Unser Ziel ist es nun, auch für den Fall des Axioms $\sim$P zu einem projektiven Raum zu gelangen, indem wir den zugehörigen

metrischen Raum in einen „Idealraum" einbetten, welcher dann die Inzidenzaxiome eines dreidimensionalen projektiven Raumes erfüllt. Die Grundlage dieser Theorie bildet das Studium zweier Relationen, welche zwischen Ebenen des metrischen Raumes bestehen können, der dreistelligen Relation des *„Im-Büschel-Liegens"* und der vierstelligen Relation des *„Im-Bündel-Liegens"*. Wir setzen in diesem Paragraphen durchweg Axiom ∼P voraus und definieren zunächst:

Definition. *Drei Ebenen α, β, γ liegen im Büschel, wenn das Produkt $\alpha\beta\gamma$ eine Erzeugende ist. Es sei für diesen Sachverhalt die abkürzende Schreibweise $[\alpha\beta\gamma]$ eingeführt.*

Ist $\alpha \neq \beta$, und haben α und β eine Schnittgerade g oder eine gemeinsame Senkrechte h, so ist die Gesamtheit der Ebenen, welche mit α, β im Büschel liegen, gleich der Gesamtheit der mit g inzidierenden bzw. der auf h senkrecht stehenden Ebenen (22′).

Durch jeden Punkt kann man zu zwei beliebigen Ebenen eine gemeinsame senkrechte Ebene finden (20). Wir zeigen nun, daß auch drei Ebenen, welche im Büschel liegen, durch senkrechte Ebenen verbindbar sind:

Satz 38 (A∼P). *Aus $\alpha \neq \gamma$, $[\alpha\beta\gamma]$; $\alpha, \gamma \perp \omega$ folgt $\beta \perp \omega$.*

Beweis. Es sei $P \mathrel{I} \beta$, $P \mathrel{\bar{I}} \alpha$, (wir können $\alpha \neq \beta$ annehmen), l das Lot durch P auf ω, $Q = l\omega$. Die Lote durch Q auf α, γ seien a, c, ihre Fußpunkte seien A, C. Ist δ die Lotebene durch Q auf die Verbindungsgerade von A, C, so folgt wie in (26) nach dem Lotensatz, daß α, $a\delta c = \beta'$, γ im Büschel liegen. Die Konstruktion ergibt außerdem (vgl. (26)), daß β' mit dem Lot durch Q auf einer Verbindungsebene von a, c inzidiert. Wegen $a, c \mathrel{I} \omega$ (16′) kann für dieses Lot die Gerade l genommen werden. Damit gilt $\beta' \mathrel{I} P$, $\beta' \perp \omega$ (9), und eine Anwendung von (34) ergibt $\beta' = \beta$, d.h. aber, es ist $\beta \perp \omega$.

Satz 39 (A∼P). *Die Relation des Im-Büschel-Liegens dreier Ebenen ist „reflexiv", „symmetrisch" und „transitiv":*

a) Aus $\alpha = \beta$ oder $\beta = \gamma$ oder $\gamma = \alpha$ folgt $[\alpha\beta\gamma]$.

b) Gilt $[\alpha\beta\gamma]$, so besteht die Büschelrelation auch für jede Permutation von α, β, γ.

c) Aus $\alpha \neq \beta$, $[\alpha\beta\gamma]$, $[\alpha\beta\delta]$ folgt $[\alpha\gamma\delta]$.

Beweis. a) $\alpha\alpha\gamma = \gamma$, $\alpha\beta\beta = \alpha$ und $\alpha\beta\alpha = \beta^{\alpha}$ liegen im Erzeugendensystem.

b) Es ist $\alpha\beta\gamma = \gamma\beta\alpha$, $\beta\gamma\alpha = (\alpha\beta\gamma)^{\alpha} = \alpha\gamma\beta$, $\gamma\alpha\beta = (\alpha\beta\gamma)^{\gamma} = \beta\alpha\gamma$.

c) Eine Lotebene ω auf α und β (20) ist nach (38) auch senkrecht zu γ und δ. Damit liegen $\alpha, \beta, \gamma, \delta$ in einer Untergruppe $\mathfrak{E}$ (37), und wir können den Beweis des Transitivitätssatzes aus der Theorie der metrischen Ebenen [*1*, § 4, Satz 6], welcher durch sechsmalige Anwendung des Lotensatzes geführt wird, übernehmen.

Aus (39) folgt nun der allgemeine Satz von den drei Spiegelungen:

Satz 40 (A∼P). *Aus $\alpha' \neq \beta'$, $[\alpha'\beta'\alpha]$, $[\alpha'\beta'\beta]$, $[\alpha'\beta'\gamma]$ folgt $[\alpha\beta\gamma]$.*

(*) *Aus $\alpha \neq \beta$, $[\alpha'\beta'\alpha]$, $[\alpha'\beta'\beta]$, $[\alpha\beta\gamma]$ folgt $[\alpha'\beta'\gamma]$.*

Beweis. Es ist $\alpha \neq \alpha'$ oder $\alpha \neq \beta'$, wir nehmen etwa $\alpha \neq \alpha'$ an. Aus $[\alpha'\beta'\alpha]$, $[\alpha'\beta'\beta]$ folgt $[\alpha'\alpha\beta]$ (39c), und aus $[\alpha'\beta'\alpha]$, $[\alpha'\beta'\gamma]$ folgt $[\alpha'\alpha\gamma]$. Wegen der Symmetrie (39b) gilt damit auch $[\alpha\alpha'\beta]$, $[\alpha\alpha'\gamma]$, und eine dritte Anwendung des Transitivitätsgesetzes ergibt $[\alpha\beta\gamma]$.

(*) Für $\alpha'=\beta'$ folgt die Behauptung aus der Reflexivität (39a). Aus $\alpha' \neq \beta'$ und $[\alpha'\beta'\alpha]$, $[\alpha'\beta'\beta]$ erhalten wir $[\alpha'\alpha\beta]$, und $[\beta'\alpha\beta]$. Zusammen mit $[\alpha\beta\gamma]$ und $\alpha \neq \beta$ ergibt sich daraus nach dem ersten Teil $[\alpha'\beta'\gamma]$.

Definition. *Die Gesamtheit der Ebenen, welche mit zwei verschiedenen Ebenen α, β im Büschel liegen, heiße ein Ebenenbüschel $g(\alpha\beta)$.*

Mit dieser Definition folgt aus Satz 40 unmittelbar der

Satz 41 (A∼P). a) *Aus $\alpha \neq \beta$ und $\alpha, \beta \in g(\alpha'\beta')$ folgt $g(\alpha'\beta') = g(\alpha\beta)$.*
b) *Aus $\alpha \neq \beta$ und $\alpha, \beta \in g(\alpha'\beta'), g(\alpha''\beta'')$ folgt $g(\alpha'\beta') = g(\alpha''\beta'')$.*

2. Ebenenbündel.

Definition. *Vier Ebenen $\alpha, \beta, \gamma, \delta$ liegen im Bündel, wenn das Produkt $\alpha\beta\gamma\delta$ einem Produkt $\xi\eta$ von zwei Erzeugenden gleich ist. Es sei für diesen Sachverhalt die abkürzende Schreibweise $[\alpha\beta\gamma\delta]$ eingeführt.*

Wir zeigen, daß die Gesamtheit der Ebenen, die mit drei durch einen Punkt P gehenden und nicht im Büschel liegenden Ebenen im Bündel liegen, gleich der Gesamtheit der mit P inzidierenden Ebenen ist:

Satz 42 (A∼P). *Satz von den vier Spiegelungen mit Umkehrung: Gilt $\alpha, \beta, \gamma, \delta \mathrm{I} P$, so folgt $[\alpha\beta\gamma\delta]$.*

(*) *Aus $\alpha, \beta, \gamma \mathrm{I} P$ und $[\alpha\beta\gamma\delta]$ folgt $\delta \mathrm{I} P$ oder $[\alpha\beta\gamma]$.*

Beweis. Der erste Teil ist gleichbedeutend mit Satz (27).

(*) Es sei $\alpha\beta\gamma\delta = \xi\eta$. Das Produkt $\xi\eta$ läßt sich durch ein $\xi'\eta'$ mit $\xi' \mathrm{I} P$ ersetzen (26), das Produkt $\eta'\delta$ sodann durch ein $\eta''\delta'$ mit $\eta'' \mathrm{I} P$. Nun ist $\alpha\beta\gamma = \xi\eta\delta = \xi'\eta'\delta = \xi'\eta''\delta'$. Da P mit $\alpha, \beta, \gamma, \xi', \eta''$ in Strichrelation steht, folgt $\delta' P = P\delta'$ also $P \mathrm{I} \delta'$ (A∼P). Aus $\eta'\delta = \eta''\delta'$ und $\eta'' = \delta'$ ergibt sich $\alpha\beta\gamma = \xi'$ also $[\alpha\beta\gamma]$. Gilt aber $\eta'' \neq \delta'$, so ergibt (22a*), daß $\delta \mathrm{I} P$ ist.

Zur Behandlung beliebiger Ebenenbündel führen wir einen Hilfsbegriff ein, der es gestattet, Mengen von Ebenen von Punkten des metrischen Raumes aus gleichsam anzupeilen:

Definition. *Es sei ein Punkt O und eine Menge von Ebenen $\mathfrak{M}$ gegeben. Zu jedem Paar $\iota, \varkappa$ aus $\mathfrak{M}$ sei ein $\omega_{\iota\varkappa}$ mit $\omega_{\iota\varkappa} \mathrm{I} O$ und $[\iota\varkappa\omega_{\iota\varkappa}]$ konstruiert. Dann heiße die Gesamtheit der $\omega_{\iota\varkappa}$ ein O-System zu $\mathfrak{M}$. Inzidieren dabei sämtliche $\omega_{\iota\varkappa}$ mit einer Geraden g, so heiße die Gesamtheit der $\omega_{\iota\varkappa}$ auch ein O-g-System zu $\mathfrak{M}$.*

Der Zusammenhang dieser Definition mit der Relation des Im-Bündel-Liegens wird hergestellt durch den

Satz 43 (A∼P). *Ist O ein beliebiger Punkt, so ist die Existenz eines O-g-Systems zu einer Menge von vier Ebenen $\alpha, \beta, \gamma, \delta$ äquivalent mit dem Bestehen der Relation $[\alpha\beta\gamma\delta]$.*

Der Beweis dieses Satzes verläuft über eine Reihe von Hilfssätzen:

Hilfssatz 1. *Zu O, α, β, γ gibt es stets ein O-g-System. Sind α, β, γ verschieden, und inzidiert O mit höchstens zwei der Ebenen α, β, γ, so ist jedes O-System zu α, β, γ ein O-g-System, liegen α, β, γ dazu nicht im Büschel, so ist g eindeutig bestimmt.*

Beweis. a, b, c seien die Lote durch O auf α, β, γ mit den Fußpunkten A, B, C. In einer Ebene ω mit $\omega I A, B, C$ mögen Verbindungsgeraden a, b, c liegen ($a I B, C$; $b I A, C$; $c I A, B$). Nach dem Lotensatz findet man $\omega_{\beta\gamma}, \omega_{\gamma\alpha}, \omega_{\alpha\beta}$ mit $\omega_{\beta\gamma}\, \omega_{\gamma\alpha}\, \omega_{\alpha\beta} = (b\alpha' c)\,(c\beta' a)\,(a\gamma' b) = (\alpha'\beta'\gamma')^b$. Aus $\alpha', \beta', \gamma' \perp \omega$ (9) folgt, daß α', β', γ' mit dem Lot durch O auf ω inzidieren (16'), und (22a) ergibt $[\alpha'\beta'\gamma']$. Damit liegen auch $\omega_{\beta\gamma}, \omega_{\gamma\alpha}, \omega_{\alpha\beta}$ im Büschel und inzidieren folglich mit einer Geraden g durch O (22'a*).

Sind α, β, γ verschieden, und inzidiert O mit höchstens einer der Ebenen α, β, γ, so sind $\omega_{\alpha\beta}, \omega_{\beta\gamma}, \omega_{\gamma\alpha}$ eindeutig bestimmt (34). Die Trägergerade g des O-g-Systems ist nur dann nicht eindeutig festgelegt, wenn $\omega_{\alpha\beta} = \omega_{\beta\gamma} = \omega_{\gamma\alpha} = \omega$ ist. Wegen $\omega \neq \alpha$ oder $\omega \neq \beta$ folgt dann aber nach (39c) aus $[\alpha\beta\omega]$, $[\beta\gamma\omega]$, $[\gamma\alpha\omega]$, daß α, β, γ im Büschel liegen.

Sind α, β, γ verschieden, und inzidiert O etwa mit α und β aber nicht mit γ, so gilt $\omega_{\beta\gamma} = \beta$, $\omega_{\gamma\alpha} = \alpha$ (34), und jedes mögliche $\omega_{\alpha\beta}$ inzidiert nach (22a*) mit der Schnittgeraden g von α und β.

Hilfssatz 2. *Es sei ein O-g-System zu α, β, γ gegeben. Ist $\alpha \neq \beta$ und ε eine Ebene mit $[\alpha\beta\varepsilon]$, so gibt es auch zu $\alpha, \beta, \gamma, \varepsilon$ ein O-g-System mit denselben g, O.*

Beweis. Nach Voraussetzung gilt $\omega_{\alpha\beta}, \omega_{\alpha\gamma}, \omega_{\beta\gamma} I g$. Diese Ebenen werden beibehalten. Aus $[\alpha\beta\omega_{\alpha\beta}]$, $[\alpha\beta\varepsilon]$ folgt $[\alpha\varepsilon\omega_{\alpha\beta}]$, $[\beta\varepsilon\omega_{\alpha\beta}]$ (39c), es sei daher $\omega_{\alpha\varepsilon} = \omega_{\beta\varepsilon} = \omega_{\alpha\beta}$. Damit ist nur noch ein $\omega_{\gamma\varepsilon}$ durch g zu finden:

Wir können voraussetzen, daß $\alpha, \beta, \gamma, \varepsilon$ verschieden sind (für $\varepsilon = \alpha$ oder $\varepsilon = \beta$ oder $\varepsilon = \gamma$ ist nichts zu zeigen, gilt $\gamma = \alpha$ oder $\gamma = \beta$, so setzen wir $\omega_{\varepsilon\gamma} = \omega_{\alpha\beta}$). Ferner sei $\gamma I O$ oder $\varepsilon I O$ (aus $\gamma, \varepsilon I O$ folgt, daß sich γ, ε in einer Geraden durch O schneiden, die wir mit g durch ein $\omega_{\gamma\varepsilon}$ verbinden können). Damit ist $\omega_{\gamma\varepsilon}$ eindeutig festgelegt (34), und (H1) ergibt, daß $[\omega_{\beta\varepsilon}\, \omega_{\gamma\varepsilon}\, \omega_{\beta\gamma}]$ und damit $[\omega_{\alpha\beta}\, \omega_{\beta\gamma}\, \omega_{\gamma\varepsilon}]$ gilt.

1. Fall: $\omega_{\alpha\beta} \neq \omega_{\beta\gamma}$. Dann inzidiert $\omega_{\gamma\varepsilon}$ mit der Schnittgeraden g von $\omega_{\alpha\beta}, \omega_{\beta\gamma}$.

2. Fall: $\omega_{\alpha\beta} \neq \omega_{\alpha\gamma}$. (H1) auf $\alpha, \gamma, \varepsilon$ angewandt liefert $[\omega_{\alpha\varepsilon}\, \omega_{\alpha\gamma}\, \omega_{\gamma\varepsilon}]$ also $[\omega_{\alpha\beta}\, \omega_{\alpha\gamma}\, \omega_{\gamma\varepsilon}]$, und $\omega_{\gamma\varepsilon}$ inzidiert mit der Schnittgeraden g von $\omega_{\alpha\beta}, \omega_{\alpha\gamma}$.

3. Fall: $\omega_{\alpha\beta} = \omega_{\alpha\gamma} = \omega_{\beta\gamma} = \omega$. Dann folgt $[\alpha\beta\gamma]$ (vgl. H1) und damit $\omega_{\gamma\varepsilon} = \omega$.

Hilfssatz 3. *Gibt es zu $\alpha, \beta, \gamma, \delta$ ein O-g-System, und ist $\alpha \neq \beta$, $\alpha\beta = \alpha'\beta'$, so gibt es zu $\alpha', \beta', \gamma, \delta$ ein O-g-System mit denselben g, O.*

Beweis. Das vorausgesetzte O-g-System enthält O-g-Systeme für α, β, γ und α, β, δ. Diese lassen sich wegen $\alpha', \beta' \in g(\alpha\beta)$ zu O-g-Systemen zu $\alpha', \alpha, \beta, \gamma$; $\beta', \alpha, \beta, \gamma$; $\alpha', \alpha, \beta, \delta$; $\beta', \alpha, \beta, \delta$ ergänzen (H2). Aus diesen vier Systemen läßt sich mit $\omega_{\alpha'\beta'} = \omega_{\alpha\beta}$ ein O-g-System zu $\alpha', \beta', \gamma, \delta$ zusammensetzen.

Hilfssatz 4. *Gilt* $[\alpha\beta\gamma]$ *oder inzidieren* $\alpha, \beta, \gamma, \delta$ *mit einem Punkt* P, *so gibt es für jedes* O *ein* O-g-*System zu* $\alpha, \beta, \gamma, \delta$.

Beweis. Wegen (H1) können wir $\alpha, \beta, \gamma, \delta$ als verschieden voraussetzen. Gilt $[\alpha\beta\gamma]$, so kann ein O-g-System zu α, β, δ (H1) nach (H2) zu einem O-g-System zu $\alpha, \beta, \gamma, \delta$ erweitert werden. Gilt $\alpha, \beta, \gamma, \delta \mathrm{I} P$, so erhält man die gesuchten $\omega_{\iota\varkappa}$ durch Verbindung der Schnittgeraden von $\iota, \varkappa$ mit einer Verbindungsgeraden von O und P.

Mit den Hilfssätzen 1, 2, 3, 4 beweisen wir nun leicht den Satz 43:

Ein Produkt $\alpha\beta\gamma\delta$ läßt sich nach (26) in folgender Weise umformen: $\alpha\beta\gamma\delta = \alpha'\beta'\gamma\delta = \alpha'\beta''\gamma'\delta = \alpha'\beta''\gamma''\delta'$ mit $\alpha', \beta'', \gamma'' \mathrm{I} O$.

a) Aus $[\alpha\beta\gamma\delta]$ folgt $[\alpha'\beta''\gamma''\delta']$ und damit $[\alpha'\beta''\gamma'']$ oder $\delta' \mathrm{I} O$ (42*). In beiden Fällen gibt es nach (H4) für jedes O ein O-g-System zu $\alpha', \beta'', \gamma'', \delta'$. Nun lehrt (H3) (bzw. im Falle $\gamma'' = \delta'$ (H1)), daß auch zu $\alpha', \beta'', \gamma', \delta$ ein O-g-System existiert. In dieser Weise kann man weiterschließen und erhält schließlich, daß auch $\alpha, \beta, \gamma, \delta$ ein O-g-System besitzen.

b) Zu $\alpha, \beta, \gamma, \delta$ existiere ein O-g-System. Dieselbe Schlußweise wie unter a) ergibt, daß auch $\alpha', \beta'', \gamma'', \delta'$ ein O-g-System besitzen. Gilt $\delta' \mathrm{I} O$, so folgt $[\alpha'\beta''\gamma''\delta']$ und damit $[\alpha\beta\gamma\delta]$ aus (42). Im anderen Falle ist δ' von $\alpha', \beta'', \gamma''$ verschieden, und (34) ergibt $\omega_{\alpha'\delta'} = \alpha'$, $\omega_{\beta''\delta'} = \beta''$, $\omega_{\gamma''\delta'} = \gamma''$ also $\alpha', \beta'', \gamma'' \mathrm{I} g$. Mit (22a) folgt dann $\alpha'\beta''\gamma'' = \xi$ und damit $\alpha\beta\gamma\delta = \alpha'\beta''\gamma''\delta' = \xi\delta'$, d.h. $[\alpha\beta\gamma\delta]$.

Satz 43 stellt ein Kriterium für das Im-Bündel-Liegen von Ebenen dar, mit dessen Hilfe es nun gelingt, auch für die Bündelrelation übersichtliche Gesetze aufzustellen.

Satz 44 (A∼P). *Die Relation des Im-Bündel-Liegens von vier Ebenen ist „reflexiv", „symmetrisch" und „transitiv":*

a) *Liegen irgend drei der Ebenen* $\alpha, \beta, \gamma, \delta$ *im Büschel, so folgt* $[\alpha\beta\gamma\delta]$.

b) *Gilt* $[\alpha\beta\gamma\delta]$, *so besteht die Bündelrelation auch für jede Permutation von* $\alpha, \beta, \gamma, \delta$.

c) *Liegen* α, β, γ *nicht im Büschel und gilt* $[\alpha\beta\gamma\delta]$, $[\alpha\beta\gamma\varepsilon]$, *so folgt* $[\alpha\beta\delta\varepsilon]$.

Beweis. a) folgt aus den Identitäten $\alpha\beta\gamma\delta = \alpha(\beta\gamma\delta) = \beta^{\alpha}(\alpha\gamma\delta) = (\alpha\beta\delta)\gamma^{\delta} = (\alpha\beta\gamma)\delta$ (oder aus (H4), (43)). b) folgt mit (43) daraus, daß die Konstruktion eines O-g-Systems nicht von der Reihenfolge der betrachteten Ebenen abhängt. Zum Beweis von c) suchen wir ein O mit $O \mathrm{I} \alpha, \beta, \gamma$, so folgt $\varepsilon \mathrm{I} O$ und damit $[\alpha\beta\delta\varepsilon]$ aus (42). Ist etwa $O \not\mathrm{I} \alpha$, so ist die Trägergerade g eines O-g-Systems zu α, β, γ eindeutig bestimmt (H1). Die nach (43) existierenden O-g-Systeme zu $\alpha, \beta, \gamma, \delta$ und zu $\alpha, \beta, \gamma, \varepsilon$ haben daher denselben Träger g, welcher wegen $O \mathrm{I} \delta$, $\omega_{\alpha\delta} = \delta$ (34) mit δ inzidiert. Somit können wir $\omega_{\delta\varepsilon} = \delta$ setzen und haben dadurch ein vollständiges O-g-System zu $\alpha, \beta, \delta, \varepsilon$ erhalten, und (43) ergibt, daß $[\alpha\beta\delta\varepsilon]$ gilt.

Aus (44) folgt nun der allgemeine Satz von den vier Spiegelungen:

Satz 45 (A∼P). *Liegen* α', β', γ' *nicht im Büschel und gilt* $[\alpha'\beta'\gamma'\alpha]$, $[\alpha'\beta'\gamma'\beta]$, $[\alpha'\beta'\gamma'\gamma]$, $[\alpha'\beta'\gamma'\delta]$, *so folgt* $[\alpha\beta\gamma\delta]$.

(*) *Liegen* α, β, γ *nicht im Büschel und gilt* $[\alpha'\beta'\gamma'\alpha]$, $[\alpha'\beta'\gamma'\beta]$, $[\alpha'\beta'\gamma'\gamma]$, $[\alpha\beta\gamma\delta]$, *so folgt* $[\alpha'\beta'\gamma'\delta]$.

Beweis. Aus (44c) folgt zunächst: $[\alpha'\beta'\alpha\beta]$, $[\alpha'\beta'\alpha\gamma]$, $[\alpha'\beta'\alpha\delta]$, $[\alpha'\beta'\beta\gamma]$, $[\alpha'\beta'\beta\delta]$, $[\alpha'\beta'\gamma\delta]$. Liegen von den Ebenen $\alpha, \beta, \gamma, \delta$ drei mit α', β' im Büschel, so ist ihr Produkt eine Erzeugende (40), und die Behauptung folgt aus (44a). Wir können daher etwa annehmen, daß α', β', α und α', β', β nicht im Büschel liegen. Dann ergibt das Transitivitätsgesetz (44c):

i) $[\alpha'\alpha\beta\gamma]$, $[\alpha'\alpha\beta\delta]$, $[\alpha'\alpha\gamma\delta]$, $[\alpha'\beta\gamma\delta]$; ii) $[\beta'\alpha\beta\gamma]$, $[\beta'\alpha\beta\delta]$, $[\beta'\alpha\gamma\delta]$, $[\beta'\beta\gamma\delta]$. Wegen $\alpha' \neq \beta'$ gilt $\alpha \neq \alpha'$ oder $\alpha \neq \beta'$. Im ersten Fall schließen wir mit i), im zweiten mit ii) weiter. Es sei etwa $\alpha \neq \alpha'$. Gilt dann $[\alpha'\alpha\beta]$, $[\alpha'\alpha\gamma]$, $[\alpha'\alpha\delta]$, so folgt $[\beta\gamma\delta]$ (40), und die Behauptung ergibt sich aus (44a). Wegen der Symmetrie von i) können wir etwa annehmen, daß α', α, β nicht im Büschel liegen. Damit folgt $[\alpha\beta\gamma\delta]$ aus $[\alpha'\alpha\beta\gamma]$, $[\alpha'\alpha\beta\delta]$ nach (44c).

(*) Gilt $[\alpha'\beta'\gamma']$, so folgt $[\alpha'\beta'\gamma'\delta]$ nach (44a). Im anderen Falle ergibt (44c): i) $[\alpha'\beta'\alpha\beta]$, $[\alpha'\beta'\alpha\gamma]$, $[\alpha'\beta'\beta\gamma]$; ii) $[\alpha'\gamma'\alpha\beta]$, $[\alpha'\gamma'\alpha\gamma]$, $[\alpha'\gamma'\beta\gamma]$. Wäre $[\alpha'\beta'\alpha]$, $[\alpha'\beta'\beta]$, $[\alpha'\beta'\gamma]$, so erhielten wir $[\alpha\beta\gamma]$ (40); liegen aber z.B. α', β', α nicht im Büschel, so folgt aus i) nach (44c), daß $[\alpha'\alpha\beta\gamma]$ und $[\beta'\alpha\beta\gamma]$ gilt. In gleicher Weise schließt man aus ii) auf $[\gamma'\alpha\beta\gamma]$. Damit haben wir $[\alpha\beta\gamma\alpha']$, $[\alpha\beta\gamma\beta']$, $[\alpha\beta\gamma\gamma']$, $[\alpha\beta\gamma\delta]$ erhalten, und der erste Teil dieses Satzes liefert $[\alpha'\beta'\gamma'\delta]$.

Definition. *Liegen drei Ebenen* α, β, γ *nicht im Büschel, so heiße die Gesamtheit der Ebenen, welche mit* α, β, γ *im Bündel liegen, ein Ebenenbündel* $P(\alpha\beta\gamma)$.

Mit dieser Definition folgt aus Satz 45 unmittelbar der

Satz 46 (A∼P). *Liegen* α, β, γ *nicht im Büschel, so gilt:*
a) *Aus* $\alpha, \beta, \gamma \in P(\alpha'\beta'\gamma')$ *folgt* $P(\alpha'\beta'\gamma') = P(\alpha\beta\gamma)$.
b) *Aus* $\alpha, \beta, \gamma \in P(\alpha'\beta'\gamma')$, $P(\alpha''\beta''\gamma'')$ *folgt* $P(\alpha'\beta'\gamma') = P(\alpha''\beta''\gamma'')$.

Die Sätze 44, 45, 46 für Bündel entsprechen den Sätzen 39, 40, 41 für Büschel. Sie stellen gleichzeitig Beziehungen zwischen beiden Relationen dar. So besagt (46b) auch, daß drei Ebenen, welche zwei verschiedenen Bündeln angehören, im Büschel liegen. Eine weitere Beziehung zwischen Büscheln und Bündeln ist der folgende

Satz 47 (A∼P). *Aus* $\varepsilon \in g(\alpha\beta)$; $\alpha, \beta \in P(\alpha'\beta'\gamma')$ *folgt* $\varepsilon \in P(\alpha'\beta'\gamma')$.

Beweis. Da α', β', γ' nicht im Büschel liegen, liegt mindestens eine dieser Ebenen nicht in $g(\alpha\beta)$ (40). Liegt etwa α' nicht mit α, β im Büschel, so folgt $P(\alpha'\beta'\gamma') = P(\alpha\beta\alpha')$ (46a). Wegen $[\varepsilon\alpha\beta\alpha']$ (44a) gilt damit $\varepsilon \in P(\alpha'\beta'\gamma')$.

Die mit einem festen Punkt inzidierenden Ebenen bilden nach Satz (42) ein Bündel. Ein Analogon dazu ist der

Satz 48 (A∼P). *Die zu einer festen Ebene* ε *senkrechten Ebenen bilden ein Bündel, das „Lotbündel"* $L(\varepsilon)$.

Beweis. Es seien vier verschiedene (44a) Ebenen $\alpha, \beta, \gamma, \delta$ mit $\alpha, \beta, \gamma, \delta \perp \varepsilon$ gegeben sowie ein Punkt O. Sämtliche Ebenen eines O-Systems zu $\alpha, \beta, \gamma, \delta$

stehen nach (38) auf ε senkrecht und inzidieren daher nach (16') mit dem Lot g durch O auf ε. Damit existiert ein O-g-System zu $\alpha, \beta, \gamma, \delta$, und (43) ergibt $[\alpha\beta\gamma\delta]$.

§ 6. Erweiterung des metrischen Raumes zum Idealraum

In diesem Paragraphen wird der zu einer (durch das Axiomensystem in § 1 gekennzeichneten) Bewegungsgruppe $\mathfrak{G}$ gehörige metrische Raum in einen dreidimensionalen projektiven Raum aus *„Idealpunkten"*, *„Idealgeraden"* und *„Idealebenen"* eingebettet. Da unter Voraussetzung von Axiom P der zu $\mathfrak{G}$ gehörige metrische Raum selbst die projektiven Inzidenzaxiome erfüllt (33), beschränken wir uns zunächst wieder auf den Fall des Axioms $\sim$P.

Als *Idealgeraden* $\boldsymbol{g}$ bezeichnen wir die Ebenenbüschel, als *Idealpunkte* $\boldsymbol{P}$ die Ebenenbündel. Den Geraden g des metrischen Raumes entsprechen nach Satz 22' eineindeutig Idealgeraden $\boldsymbol{g} = g(\alpha\beta)$ mit $\alpha, \beta \mathrm{I} g$; den Punkten P des metrischen Raumes entsprechen nach Satz 42 eineindeutig Idealpunkte $\boldsymbol{P} = P(\alpha\beta\gamma)$ mit $\alpha, \beta, \gamma \mathrm{I} P$. Solche *„eigentlichen"* Idealgeraden und Idealpunkte werden wir in der Bezeichnungsweise nicht von den entsprechenden Geraden und Punkten des metrischen Raumes unterscheiden. Die folgenden Inzidenzdefinitionen sind so eingerichtet, daß eigentliche Idealelemente genau dann miteinander inzidieren, wenn die ihnen zugeordneten Elemente des metrischen Raumes im bisherigen Sinne inzidieren (22') (42):

Definition. $\varepsilon \mathrm{I} g(\alpha\beta)$ *bedeute* $\varepsilon \in g(\alpha\beta)$; $\varepsilon \mathrm{I} P(\alpha\beta\gamma)$ *bedeute* $\varepsilon \in P(\alpha\beta\gamma)$; $g(\xi\eta) \mathrm{I} P(\alpha\beta\gamma)$ *bedeute* $\xi, \eta \in P(\alpha\beta\gamma)$.

Aus diesen Definitionen und aus den Sätzen über Büschel und Bündel ergeben sich nun leicht eine Reihe von Aussagen über eigentliche (P, g, ε) und nicht-(notwendig)-eigentliche $(\boldsymbol{P}, \boldsymbol{g})$ Idealelemente:

i) Aus $\boldsymbol{P} \mathrm{I} \boldsymbol{g}$ und $\boldsymbol{g} \mathrm{I} \varepsilon$ folgt $\boldsymbol{P} \mathrm{I} \varepsilon$ ((i) folgt aus (47)).

ii) Zu $P, \boldsymbol{g}$ gibt es ein ε mit $\varepsilon \mathrm{I} P, \boldsymbol{g}$ ((ii) folgt aus (26)).

iii) Zu $P, \boldsymbol{Q}$ gibt es ein g mit $g \mathrm{I} P, \boldsymbol{Q}$.

(Beweis. Es sei $\boldsymbol{Q} = P(\alpha\beta\gamma)$; $\xi, \eta, \zeta \mathrm{I} P$; $\xi \mathrm{I} g(\alpha\beta)$, $\eta \mathrm{I} g(\beta\gamma)$, $\zeta \mathrm{I} g(\gamma\alpha)$ (ii). $\xi = \eta = \zeta$ ist ausgeschlossen, da dann $[\alpha\beta\gamma]$ folgte (39). Ist nun etwa $\xi \neq \eta$, so gilt $g(\xi\eta) \mathrm{I} P, \boldsymbol{Q}$; und da nach (11 b) eine Schnittgerade von ξ, η existiert, ist $g(\xi\eta)$ eigentlich.)

iv) Zu $P, \boldsymbol{Q}, \boldsymbol{R}$ gibt es ein ε mit $\varepsilon \mathrm{I} P, \boldsymbol{Q}, \boldsymbol{R}$.

(Beweis. Zu $P, \boldsymbol{Q}$ gibt es ein g mit $g \mathrm{I} P, \boldsymbol{Q}$ (iii), zu $P, \boldsymbol{R}$ ein h mit $h \mathrm{I} P, \boldsymbol{R}$. Für eine Verbindungsebene ε von g, h (12c) folgt dann nach (i): $\varepsilon \mathrm{I} P, \boldsymbol{Q}, \boldsymbol{R}$.)

Mit diesen Vorbemerkungen beweisen wir nun den

Satz 49 (A$\sim$P). *Zu zwei verschiedenen Idealpunkten* $\boldsymbol{P}, \boldsymbol{Q}$ *gibt es genau eine Idealgerade, die mit* $\boldsymbol{P}$ *und* $\boldsymbol{Q}$ *inzidiert.*

Beweis. Es sei A ein beliebiger Punkt, α eine Ebene mit $\alpha \mathrm{I} A, \boldsymbol{P}, \boldsymbol{Q}$ (iv), B ein Punkt mit $B \nmid \alpha$, und β eine Ebene mit $\beta \mathrm{I} B, \boldsymbol{P}, \boldsymbol{Q}$ (iv). Dann gilt $g(\alpha\beta) \mathrm{I} \boldsymbol{P}, \boldsymbol{Q}$. Drei Ebenen, welche zwei verschiedenen Bündeln angehören, liegen nach (46b) im Büschel, damit ist die ideale Verbindungsgerade von $\boldsymbol{P}$ und $\boldsymbol{Q}$ eindeutig bestimmt.

Satz 50 (A∼P). *Es sei* $\boldsymbol{g} = g(\alpha\beta)$, $\boldsymbol{h} = g(\gamma\delta)$, $\boldsymbol{g} \neq \boldsymbol{h}$. *Liegen* $\alpha, \beta, \gamma, \delta$ *im Bündel, so inzidieren* $\boldsymbol{g}$ *und* $\boldsymbol{h}$ *mit einem Idealpunkt* $\boldsymbol{P}$. *Sind* $\boldsymbol{g}, \boldsymbol{h}$ *eigentliche Idealgeraden, so inzidieren* $\boldsymbol{g}, \boldsymbol{h}$ *außerdem mit einer Ebene* ε.

Beweis. Liegen α, β, γ bzw. α, β, δ nicht im Büschel, so folgt aus $[\alpha\beta\gamma\delta]$ sofort $\boldsymbol{g}, \boldsymbol{h} \mathrm{I} P(\alpha\beta\gamma)$ bzw. $\boldsymbol{g}, \boldsymbol{h} \mathrm{I} P(\alpha\beta\delta)$. Gilt aber $[\alpha\beta\gamma]$ und $[\alpha\beta\delta]$, so sind $\boldsymbol{g}$ und $\boldsymbol{h}$ gleich (41a). Sind $\boldsymbol{g}, \boldsymbol{h}$ eigentlich, so verbinden wir einen Punkt Q auf $\boldsymbol{g}$ ($Q \neq \boldsymbol{P}$) mit $\boldsymbol{h}$ durch ein ε (12b), welches dann auch mit $\boldsymbol{P}$ inzidiert (i). Da ε den zu $\boldsymbol{P}$ und Q gehörigen Bündeln angehört, folgt nach (46b) $[\alpha\beta\varepsilon]$, d.h. ε inzidiert auch mit $\boldsymbol{g}$.

Die weitere Durchführung der Erweiterung des metrischen Raumes schließen wir an die Arbeit [9] von A. WINTERNITZ an. Dort wird untersucht, wann ein Raum $\mathfrak{H}$, in welchem die Hilbertschen Inzidenzaxiome erfüllt sind, zu einem projektiven Raum ergänzt werden kann. Wir zeigen zunächst, daß unser metrischer Raum ein solcher Raum $\mathfrak{H}$ ist:

Das Erzeugendensystem $\mathfrak{S}$ von $\mathfrak{G}$ ist nicht leer, daher gibt es eine Ebene α und auf dieser einen Punkt $P \equiv (\alpha\beta\gamma)$ (A6) (3). Nach (6) inzidiert mit den Geraden $(\alpha\beta)$, $(\beta\gamma)$, $(\gamma\alpha)$ je ein von P verschiedener Punkt. Man sieht leicht ein, daß damit die absoluten Existenzforderungen I_8 und I_3, 2. Teil (Bezeichnung nach [3]) erfüllt sind. Der erste Teil von I_3 wird von WINTERNITZ in der verschärften Fassung angewandt, daß mit jeder Geraden mindestens drei Punkte inzidieren, dieses ist gleichbedeutend mit (6). I_1 und I_2 sind äquivalent zu (10), I_4 und I_5 folgen aus (12a) und (A6), I_6 ergibt sich aus (11a) und (8) und I_7 schließlich ist gleichbedeutend mit (A5).

Hinreichend für die Ergänzbarkeit eines Raumes $\mathfrak{H}$ zu einem projektiven Raum sind nun nach WINTERNITZ die folgenden Hypothesen:

H1) *(Vierkanthypothese). Haben zwei Dreikante zwei Kanten gemein, so sind die dritten Kanten in einer Ebene gelegen.*

(Unter einem Dreikant wird hierbei eine Figur von drei Ebenen verstanden, welche keine Gerade (höchstens einen Punkt) gemein haben, aber sich paarweise schneiden. Die drei Schnittgeraden („Kanten") liegen nicht in einer Ebene, aber je zwei sind komplanar).

H2) *Die Gültigkeit des Desarguesschen Satzes im Bündel.*

Auch diese beiden Hypothesen sind im metrischen Raum erfüllt:

Beweis von H1. Es seien α, β, γ und α', β', γ die Ebenen der betrachteten Dreikante. Die gemeinsamen Kanten a und b entsprechen den Büscheln $g(\beta\gamma) = g(\beta'\gamma)$ und $g(\alpha\gamma) = g(\alpha'\gamma)$. Nach (49) ist der ideale Schnittpunkt $P(\alpha\beta\gamma)$ bzw. $P(\alpha'\beta'\gamma)$ der verschiedenen Geraden a und b eindeutig bestimmt, daher inzidieren mit ihm sämtliche fünf Ebenen (i), insbesondere liegen $\alpha, \beta, \alpha', \beta'$ im Bündel. Nach (50) sind daher die dritten (eigentlichen) Kanten $g(\alpha\beta)$ und $g(\alpha'\beta')$ der beiden Dreikante komplanar.

Beweis von H2. Der Satz von DESARGUES im Bündel kann mit Hilfe der schon gezeigten Sätze über Idealelemente durch eine Modifikation des (nicht anordnungsfreien) Beweises bei F. SCHUR [6] gewonnen werden. Kürzer sieht man die Gültigkeit dieses Satzes folgendermaßen ein: Die Gesamtheit der Ebenen und Geraden durch einen festen Punkt bildet ein Modell für die

Gruppenebene einer ebenen elliptischen Bewegungsgruppe (§ 4). In dieser gilt der Satz von PAPPUS-PASCAL (vgl. den einfachen Beweis (ohne Gegenpaarung und ohne Halbdrehungen) in [*1*, § 16, 2]), also auch der Satz von DESARGUES.

Man sieht leicht, daß die Definitionen der Idealpunkte und Idealgeraden bei WINTERNITZ unter Voraussetzung der bisherigen Sätze dieses Paragraphen dasselbe beschreiben, wie die bei uns verwendeten Definitionen, so daß die Beweise in [*9*, III] von uns wörtlich übernommen werden können. Idealebenen werden nach [*9*] so definiert:

Definition. *Zwei Dreiecke* $\boldsymbol{A, B, C}$ *und* $\boldsymbol{A', B', C'}$ *aus nichtkollinearen Idealpunkten definieren die gleiche Idealebene* $\mathfrak{e} = (\boldsymbol{A, B, C}) = (\boldsymbol{A', B', C'})$, *wenn die (idealen) Seiten des einen Dreiecks die des anderen in Idealpunkten schneiden.*

Die so erklärte „Ebenengleichheit" von Dreiecken ist reflexiv, symmetrisch und transitiv. Dies folgt aus dem bei WINTERNITZ bewiesenen Axiom $\mathfrak{B}$, A3 von VEBLEN-YOUNG. Den Ebenen des metrischen Raumes entsprechen nach (12a) eineindeutig „eigentliche" Idealebenen, die durch eigentliche Idealpunkte definiert werden können. Die Inzidenz mit Idealebenen wird nach [*9*] in folgender Weise erklärt:

Definition. $\boldsymbol{g}\,\mathrm{I}(\boldsymbol{A, B, C})$ *sei gleichbedeutend damit, daß* $\boldsymbol{g}$ *die Seiten des Dreiecks* $\boldsymbol{A, B, C}$ *in Idealpunkten schneidet.* $\boldsymbol{P}\,\mathrm{I}\,\mathfrak{e}$ *gelte genau dann, wenn es ein* $\boldsymbol{g}$ *gibt, welches mit* $\boldsymbol{P}$ *und mit* $\mathfrak{e}$ *inzidiert.*

Das Ergebnis der Winternitzschen Arbeit ergibt nun, daß der aus den Idealpunkten, den Idealgeraden und den Idealebenen bestehende „Idealraum" ein projektiver Raum ist. Da die Idealpunkte und die Idealgeraden stets als Schnittgebilde *eigentlicher* Idealebenen darstellbar sind, trägt die hier durchgeführte Erweiterung Minimalcharakter, keines der eingeführten Idealelemente ist entbehrlich, und es gilt

Satz 51 (A∼P). a) *Aus* $P\,\mathrm{I}\,\boldsymbol{g}$ *folgt:* $\boldsymbol{g}$ *ist eigentlich.*

b) *Aus* $P\,\mathrm{I}\,\mathfrak{e}$ *oder* $g\,\mathrm{I}\,\mathfrak{e}$ *folgt:* $\mathfrak{e}$ *ist eigentlich.*

Beweis. a) folgt aus (11b). b) folgt aus (iv) und (6).

Aus der Gültigkeit des Satzes von PAPPUS-PASCAL und des Fano-Axioms in der Gruppenebene einer ebenen elliptischen Bewegungsgruppe im Sinne von BACHMANN [*1*] folgt die Gültigkeit dieser Sätze im Bündelmodell um einen eigentlichen Idealpunkt. Durch eine geeignete Projektion ergibt sich hieraus, daß der Satz von PAPPUS-PASCAL und das Fano-Axiom auch in jeder Idealebene und in jedem Bündel gelten.

Wir fassen die Ergebnisse dieses Paragraphen zusammen in dem

Satz 52. *Jeder (zu einer Gruppe* $\mathfrak{G}$ *und einem Erzeugendensystem* $\mathfrak{S}$ *gehörige) metrische Raum ist Teilraum eines projektiven Raumes, in welchem der Satz von* PAPPUS-PASCAL *und das Fano-Axiom gelten.*

§ 7. Die absolute Polarität

1. Projektiv-metrische Räume. Das Ziel dieses letzten Abschnittes ist es, zu zeigen, daß die durch das Axiomensystem in § 1 gekennzeichneten Bewegungsgruppen $\mathfrak{G}$ Untergruppen von Bewegungsgruppen projektiv-metrischer

Räume sind. Unter einem *projektiv-metrischen Raum* verstehen wir einen projektiven Raum $\mathfrak{R}$, in welchem der Satz von PAPPUS-PASCAL und das Fano-Axiom gelten und in dem es entweder a) eine projektive Polarität Γ in einer festen Ebene ω gibt mit der Eigenschaft, daß vermöge Γ zugeordnete Punkte und Geraden auf ω nicht miteinander inzidieren, oder in dem es b) eine projektive Polarität Π (für den gesamten Raum) gibt. In beiden Fällen erklären wir dann eine Orthogonalität:

a) Eine Ebene α sei senkrecht zu einer Ebene β, wenn α mit dem vermöge Γ der Schnittgeraden von ω und β zugeordneten Punkt inzidiert. ω sei außerdem zu allen Ebenen des Raumes $\mathfrak{R}$ orthogonal.

b) Eine Ebene α sei senkrecht zu einer Ebene β, wenn α mit dem vermöge Π der Ebene β zugeordneten Punkt inzidiert.

Das Senkrechtstehen in $\mathfrak{R}$ ist eine symmetrische Relation, weil Γ und Π involutorische Korrelationen sind. $\mathfrak{R}$ heiße im Falle a) ein *„singulärer"* im Falle b) ein *„ordinärer"* projektiv-metrischer Raum. Ist in einem ordinären projektiv-metrischen Raum eine harmonische Homologie Σ mit einem Paar Pol-Polarebene (P, P^{Π}) als Zentrum und Achse gegeben, so ist Σ mit Π vertauschbar. In einem singulären projektiv-metrischen Raum sind mit der Polarität Γ die harmonischen Homologien mit einer Ebene ε als Achse und dem vermöge Γ der Schnittgeraden von ε und ω zugeordneten Punkt als Zentrum vertauschbar.

Als Bewegungsgruppe eines projektiv-metrischen Raumes werde in beiden Fällen die aus den soeben betrachteten harmonischen Homologien erzeugte Gruppe von projektiven Kollineationen erklärt. Diese ist also eine Untergruppe der Gruppe aller Projektivitäten, welche die „absolute" Polarität Π bzw. Γ erhalten.

Den beiden Arten projektiv-metrischer Räume entsprechend, haben wir bei unseren Gruppen $\mathfrak{G}$ zwei Fälle zu unterscheiden. Dem Fall a) entspricht das Axiom der euklidischen Metrik (Axiom R), welches z.B. von der Bewegungsgruppe des klassischen euklidischen Raumes erfüllt wird; dem Fall b) entspricht das Axiom der nichteuklidischen Metrik (Axiom $\sim$R), welchem die Bewegungsgruppen der klassischen nichteuklidischen Räume gehorchen:

Axiom R. *Es gibt* $\pi, \vartheta, \alpha, \beta, \gamma$ *mit* $\pi \neq \vartheta$; $\alpha, \beta, \gamma \,|\, \pi, \vartheta$ *und* $\alpha\beta\gamma \notin \mathfrak{S}$.

Axiom $\sim$R. *Aus* $\pi \neq \vartheta$; $\alpha, \beta, \gamma \,|\, \pi, \vartheta$ *folgt* $\alpha\beta\gamma \in \mathfrak{S}$.

Das Axiom $\sim$R ist eine Verschärfung von Axiom 4. Der Transitivitätssatz für Ebenenbüschel (39c) folgt im Falle $\sim$R in trivialer Weise aus (38). Gilt Axiom P, so folgt Axiom $\sim$R aus Axiom 3 (Man ersetze in (A3) P, Q durch ihre Polarebenen). Axiom R zieht also stets Axiom $\sim$P nach sich.

Entsprechend dem Symbol $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ für Idealebenen bezeichnen wir im folgenden die ideale Verbindungsgerade zweier verschiedener Idealpunkte $\boldsymbol{A}, \boldsymbol{B}$ auch mit $(\boldsymbol{A}, \boldsymbol{B})$, die ideale Schnittgerade zweier verschiedener Idealebenen $\boldsymbol{\alpha}, \boldsymbol{\beta}$ mit $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ und den idealen Schnittpunkt dreier, nicht mit einer Idealgeraden inzidierender, Idealebenen $\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma$ mit $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$.

Zur Vorbereitung der Konstruktion einer absoluten Polarität Γ bzw. Π benötigen wir einen Satz über Lotbündel:

Satz 53 (A$\sim$P). a) *Gilt* $[\alpha\beta\gamma]$, *so inzidieren* $L(\alpha), L(\beta), L(\gamma)$ *mit einer Idealgeraden.*

b) *Gilt* $[\alpha\beta\gamma\delta]$, *so inzidieren* $L(\alpha), L(\beta), L(\gamma), L(\delta)$ *mit einer Idealebene.*

Beweis. a) Wir können $\alpha \neq \beta$ annehmen. Zwei verschiedene Ebenen σ, τ mit $\sigma, \tau \mid \alpha, \beta$ (20) stehen auch auf γ senkrecht (38). Damit inzidiert $g(\sigma\tau)$ mit $L(\alpha), L(\beta), L(\gamma)$.

b) Nach (a) können wir voraussetzen, daß keine drei der Ebenen $\alpha, \beta, \gamma, \delta$ im Büschel liegen. Ist C ein Punkt mit $C\,\mathrm{I}\,\gamma$, so gibt es eine Ebene ε mit $\varepsilon\,\mathrm{I}\,C$, $g(\alpha\beta)$ (26), also auch $\varepsilon\,\mathrm{I}\,P(\alpha\beta\gamma) = \boldsymbol{P}$. Nach (51a) ist $(C, \boldsymbol{P}) = g(\varepsilon\gamma)$ eine eigentliche Idealgerade, die mit $\boldsymbol{P}$ inzidiert und daher mit $g(\beta\delta)$ durch eine eigentliche (51b) Idealebene ω verbindbar ist. (Da α, β, γ nicht im Büschel liegen sollen, ist $\varepsilon \neq \alpha$ oder $\varepsilon \neq \beta$ (39a), wir nehmen $\varepsilon \neq \beta$ an). Nun steht nach (38) jede zu ε, γ und jede zu β, δ senkrechte Ebene auch auf ω senkrecht. Ideale Verbindungsgeraden von $L(\varepsilon), L(\gamma)$ und von $L(\beta), L(\delta)$ schneiden sich daher in $L(\omega)$, und aus den projektiven Inzidenzaxiomen folgt, daß $L(\varepsilon), L(\beta), L(\gamma), L(\delta)$ mit einer Idealebene $\boldsymbol{\varphi}$ inzidieren.

Ist $L(\varepsilon) = L(\beta)$, so folgt aus $\varepsilon \neq \beta$ nach (38) $L(\varepsilon) = L(\alpha)$ also $L(\alpha)\,\mathrm{I}\,\boldsymbol{\varphi}$. Gilt dagegen $L(\varepsilon) \neq L(\beta)$, so ist $L(\alpha)\,\mathrm{I}\,(L(\varepsilon), L(\beta))$ (a) und damit wieder $L(\alpha)\,\mathrm{I}\,\boldsymbol{\varphi}$.

2. Räume mit euklidischer Metrik. Wir setzen in dieser Nummer Axiom R (und damit auch Axiom $\sim$P) voraus.

Definition. *Zwei Ebenen* α, β *heißen parallel* $(\alpha \| \beta)$, *wenn ihre Lotbündel übereinstimmen (wenn* $L(\alpha) = L(\beta)$ *ist).*

Die Ebenen α, β, γ aus Axiom R definieren (da sie nicht im Büschel liegen) ein Bündel $P(\alpha\beta\gamma)$, welches gemeinsames Lotbündel zu den Ebenen π und ϑ aus Axiom R ist. Es gibt also Ebenen π, ϑ mit $\pi \neq \vartheta$ und $\pi \| \vartheta$. Ein Kriterium für das Bestehen der Parallelenrelation ist

Satz 54 (A$\sim$P). *Aus* $\pi \| \vartheta$, $g \perp \pi$ *folgt* $g \perp \vartheta$.

(*) *Aus* $g \neq h$; $g, h \perp \pi, \vartheta$ *folgt* $\pi \| \vartheta$.

Beweis. Es sei $g = \alpha\beta$. $g \perp \pi$ bedeutet $\alpha, \beta \perp \pi$, und $L(\pi) = L(\vartheta)$ ergibt $\alpha, \beta \perp \vartheta$ und damit $g \perp \vartheta$ (15b*).

(*) Ist $g \neq h$, $g = \alpha\beta$, $h = \gamma\delta$, so sind α, β, γ (oder α, β, δ) drei nicht im Büschel liegende Ebenen (41a), und aus $g, h \perp \pi, \vartheta$ folgt $P(\alpha\beta\gamma) = L(\pi) = L(\vartheta)$.

Mit Axiom R läßt sich (54*) erweitern zu dem

Satz 55 (AR). *Gilt* $s \perp \sigma, \tau$, *so folgt* $\sigma \| \tau$.

Beweis. S sei ein Punkt mit $S\,\mathrm{I}\,s$. Das Lot l durch S auf die Ebene π des Axioms R steht auch auf ϑ senkrecht (54). Daher gilt für ein δ mit $\delta\,\mathrm{I}\,l$, s nach (9) $\delta \perp \sigma, \tau, \pi, \vartheta$. Zu einem Punkt P mit $P \not\mathrm{I}\,\delta$ gibt es nach (20) ein ε mit $\varepsilon\,\mathrm{I}\,P$, $\varepsilon \perp \sigma, \pi$ also $\varepsilon \perp \sigma, \pi, \vartheta$ (π, ϑ sind parallel (AR)).

1. Fall: Es sei $[\sigma\pi\vartheta]$. Dann gilt nach (22'b*) $\sigma \perp l$ (22'b*), $l = s$ (15'a*), also $[\tau\pi\vartheta]$ (22b), und aus $L(\pi) = L(\vartheta)$ folgt mit (38) $L(\sigma) = L(\tau) = L(\pi)$.

2. Fall: σ, π, ϑ liegen nicht im Büschel. Dann inzidiert ε nicht mit s, denn aus $\varepsilon \mathrel{I} s$; $\varepsilon, \delta \mathrel{I} s$; $\varepsilon, \delta \perp \pi, \vartheta$ folgte $\pi, \vartheta \perp s$ (15b*) also $[\sigma\pi\vartheta]$ (22b). Nun ist $L(\delta) = L(\varepsilon) = P(\sigma\pi\vartheta)$, somit gilt auch $\tau \in L(\varepsilon)$ d.h. $\varepsilon \perp \tau$. Zwei verschiedene Ebenen α, β mit $\alpha, \beta \mathrel{I} s$ und die Ebene ε definieren daher ein Bündel $P(\alpha\beta\varepsilon)$, welches gemeinsames Lotbündel zu σ und τ ist.

Satz 56 (AR). *Zu P, σ gibt es genau ein τ mit $\tau \mathrel{I} P$, $\tau \parallel \sigma$.*

Beweis. Ist l das Lot durch P auf σ und τ die Lotebene in P senkrecht zu l, so gilt $\tau \parallel \sigma$ (55). Da jede zu σ parallele Ebene auf l senkrecht steht (54), folgt die Eindeutigkeit der Parallelebene durch P aus (21'a*).

Satz 57 (AR). *Aus $\alpha \not\parallel \beta$; $\delta, \varepsilon \perp \alpha, \beta$ folgt $\delta \parallel \varepsilon$.*

Beweis. Wir konstruieren durch einen Punkt P mit $P \mathrel{I} \alpha$ die Parallelebene γ zu β (56). Aus $L(\gamma) = L(\beta)$ folgt dann $\gamma \perp \delta, \varepsilon$. Die Schnittgerade von α und γ steht auf δ und ε senkrecht (15b*), daher gilt $\delta \parallel \varepsilon$ (55).

Satz 58 (AR). *Es gibt eine Idealebene $\boldsymbol{\omega}$ mit der Eigenschaft, daß die Gesamtheit der mit ihr inzidierenden Idealpunkte gleich der Gesamtheit der Lotbündel ist. Die Gesamtheit der mit $\boldsymbol{\omega}$ inzidierenden Idealgeraden ist gleich der Gesamtheit der Büschel $g(\alpha\beta)$ mit $\alpha \parallel \beta$.*

Beweis. Es seien $L(\alpha), L(\beta), L(\gamma), L(\delta)$ vier verschiedene Lotbündel. Sind nun $\alpha', \beta', \gamma', \delta'$ die Parallelebenen zu $\alpha, \beta, \gamma, \delta$ durch einen festen Punkt P (56), so folgt $L(\alpha) = L(\alpha')$ usw., und aus $[\alpha'\beta'\gamma'\delta']$ (42) ergibt sich, daß $L(\alpha)$, $L(\beta), L(\gamma), L(\delta)$ mit einer Idealebene $\boldsymbol{\omega}$ inzidieren (53b). Andererseits ist jedes $\boldsymbol{P}$ mit $\boldsymbol{P} \mathrel{I} \boldsymbol{\omega}$ ein Lotbündel: Wir verbinden $\boldsymbol{P}$ mit einem Q, welches nicht mit $\boldsymbol{\omega}$ inzidiert, durch ein g (51a). Das Lotbündel $L(\sigma)$ einer zu g senkrechten Ebene σ inzidiert mit g und mit $\boldsymbol{\omega}$, ist also der eindeutig bestimmte Schnittpunkt von g und $\boldsymbol{\omega}$, d.h. es gilt $\boldsymbol{P} = L(\sigma)$.

Eine Idealgerade auf $\boldsymbol{\omega}$ ist stets ein $(L(\alpha), L(\beta))$ mit $\alpha \not\parallel \beta$ und daher nach (57) ein Büschel paralleler Ebenen.

Auf $\boldsymbol{\omega}$ erklären wir nun die Abbildung $\boldsymbol{\Gamma}$ wie folgt: Ist $\alpha \parallel \beta$, so sei $(g(\alpha\beta))^{\boldsymbol{\Gamma}} = L(\alpha) = L(\beta)$. Umgekehrt sei $(L(\alpha))^{\boldsymbol{\Gamma}}$ das Büschel der zu α parallelen Ebenen.

Diese Abbildung ist offenbar eindeutig und involutorisch also auch umkehrbar eindeutig. Nach (58) werden von $\boldsymbol{\Gamma}$ sämtliche mit $\boldsymbol{\omega}$ inzidierenden Idealpunkte und -geraden erfaßt. Aus $g(\alpha\beta) \mathrel{I} L(\gamma)$, $\alpha \parallel \beta$, $\delta \parallel \gamma$ folgt $\alpha, \beta \perp \gamma$ also auch $\alpha, \beta \perp \delta$. Damit gilt $L(\alpha) \mathrel{I} g(\gamma\delta)$, d.h. $\boldsymbol{\Gamma}$ ist inzidenztreu. Weiter ist stets $g(\alpha\beta) \not\mathrel{I} L(\alpha)$, da sonst $\alpha \perp \alpha$ wäre. $\alpha \perp \gamma$ bedeutet, daß α mit $L(\gamma)$ inzidiert, $L(\gamma)$ ist aber durch $\boldsymbol{\Gamma}$ der idealen Schnittgeraden von α und $\boldsymbol{\omega}$ zugeordnet; $\boldsymbol{\Gamma}$ ist also eine involutorische Korrelation auf $\boldsymbol{\omega}$, welche die Orthogonalität im metrischen Raum bestimmt.

Aus der Gültigkeit des Höhensatzes in einer elliptischen Ebene (vgl. [*1*, § 4, 2]) und damit auch im Bündelmodell um einen Punkt P ergibt sich die Gültigkeit des Höhensatzes in $\boldsymbol{\omega}$ (die Höhen eines Dreiecks in $\boldsymbol{\omega}$ sind dabei die idealen Verbindungsgeraden der Ecken mit den durch $\boldsymbol{\Gamma}$ den entsprechenden Seiten zugeordneten Idealpunkten). Gilt aber in einer projektiven Ebene für eine Polarität der Höhensatz, so ist die Polarität projektiv

(H. LENZ [5] vgl. [1, § 5, 5]). Zusammenfassend erhalten wir damit durch Vergleich mit Nr. 1 den

Satz 59 (AR). *Unter Voraussetzung von Axiom R ist der Idealraum des zu $\mathfrak{G}$ gehörigen metrischen Raumes ein singulärer projektiv-metrischer Raum.*

3. Räume mit nichteuklidischer Metrik. Wir betrachten nun den Fall des Axioms ∼R. Gilt Axiom P, so gibt es im metrischen Raum, welcher schon ein projektiver Raum ist, eine die Orthogonalität vermittelnde Polarität (32) (33), wir setzen daher zusätzlich Axiom ∼P voraus und betrachten zunächst wieder Lotbündel:

Satz 60 (A∼R) (A∼P). *Aus $L(\alpha)=L(\beta)$ folgt $\alpha=\beta$.*

Beweis. Es sei $L(\alpha)=L(\beta)=P(\varrho\sigma\tau)$. Da ϱ, σ, τ nicht im Büschel liegen, folgt aus (A∼R), daß $\alpha=\beta$ ist.

Zu Satz 53 beweisen wir eine Umkehrung:

Satz 61 (A∼R) (A∼P). a) *Inzidieren $L(\alpha), L(\beta), L(\gamma)$ mit einer Idealgeraden, so gilt $[\alpha\beta\gamma]$.*

b) *Inzidieren $L(\alpha), L(\beta), L(\gamma), L(\delta)$ mit einer Idealebene, so folgt $[\alpha\beta\gamma\delta]$.*

Beweis. a) Aus $g(\sigma\tau)\,\mathrm{I}\,L(\alpha), L(\beta), L(\gamma)$, d.h. $\sigma, \tau | \alpha, \beta, \gamma$ folgt $[\alpha\beta\gamma]$ nach (A∼R).

b) Wir können annehmen, daß $\alpha, \beta, \gamma, \delta$ und damit (60) $L(\alpha), L(\beta), L(\gamma), L(\delta)$ verschieden sind (44a). Ist dann $\big(L(\alpha), L(\beta)\big)=g(\pi\varrho)$ und $\big(L(\gamma), L(\delta)\big)=g(\sigma\tau)$, so liegen auch $g(\pi\varrho)$ und $g(\sigma\tau)$ in der betrachteten Idealebene und schneiden sich daher in (mindestens) einem Idealpunkt, d.h. es gilt $[\pi\varrho\sigma\tau]$. Wegen $\pi\neq\varrho, \sigma\neq\tau$ gilt nun auch $g(\alpha\beta)=\big(L(\pi), L(\varrho)\big)$ und $g(\gamma\delta)=\big(L(\sigma), L(\tau)\big)$. Aus $[\pi\varrho\sigma\tau]$ folgt nach (53b), daß $L(\pi), L(\varrho), L(\sigma), L(\tau)$ mit einer Idealebene inzidieren. Damit schneiden sich $g(\alpha\beta)$ und $g(\gamma\delta)$ in einem Idealpunkt, d.h. $\alpha, \beta, \gamma, \delta$ liegen im Bündel.

Wir erklären nun die Abbildung $\mathbf{\Pi}$ zunächst für Idealgeraden: Ist $\boldsymbol{g}=g(\alpha\beta)$, so sei $\boldsymbol{g}^{\mathbf{\Pi}}=\boldsymbol{g}^{*}$ eine Idealgerade $g(\sigma\tau)$ mit $\sigma, \tau\perp\alpha, \beta$. Eine solche „*Polare*" von $\boldsymbol{g}$ ist mit (20) stets konstruierbar, und läßt sich als $\big(L(\alpha), L(\beta)\big)$ schreiben. Die Abbildung ist unabhängig von der Darstellung von $\boldsymbol{g}$, denn aus $g(\alpha\beta)=g(\gamma\delta)$ folgt nach (38), daß auch $\sigma, \tau\perp\gamma, \delta$ gilt. Sämtliche Ebenen ε mit $\varepsilon\perp\alpha, \beta$ inzidieren nach (A∼R) mit einer Idealgeraden, daher ist $\boldsymbol{g}^{*}$ eindeutig bestimmt. $(\boldsymbol{g}^{*})^{*}=\boldsymbol{g}^{**}$ inzidiert mit allen zu σ, τ senkrechten Ebenen, fällt also wegen $\sigma\neq\tau$ mit $\boldsymbol{g}$ zusammen (A∼R). Damit ist $\mathbf{\Pi}$ involutorisch (für Idealgeraden) also auch umkehrbar eindeutig.

Satz 62 (A∼R) (A∼P). *Schneiden sich zwei Idealgeraden in einem Idealpunkt, so schneiden sich auch deren Polaren.*

Beweis. Die Idealgeraden seien $g(\alpha\beta)$ und $g(\gamma\delta)$. Aus $[\alpha\beta\gamma\delta]$ folgt, daß $L(\alpha), L(\beta), L(\gamma), L(\delta)$ in einer Idealebene liegen (53b). Damit schneiden sich $\big(L(\alpha), L(\beta)\big)$ und $\big(L(\gamma), L(\delta)\big)$ in einem Idealpunkt.

Für Idealpunkte erklären wir die Abbildung $\mathbf{\Pi}$ in folgender Weise: Ist $\boldsymbol{P}=P(\alpha\beta\gamma)$, so sei die „Polarebene" $\boldsymbol{P}^{\mathbf{\Pi}}=\boldsymbol{P}^{*}$ von $\boldsymbol{P}$ die Idealebene $\big(L(\alpha),$

$L(\beta), L(\gamma)\big)$ (nach (61a) inzidieren $L(\alpha), L(\beta), L(\gamma)$ nicht mit einer Idealgeraden). Satz (53b) lehrt, daß die Definition von $\boldsymbol{P}^*$ nicht von der Darstellung von $\boldsymbol{P}$ (durch α, β, γ) abhängt. Damit bildet $\boldsymbol{\Pi}$ sämtliche Idealpunkte eindeutig in die Menge der Idealebenen ab.

Satz 63 (A∼R) (A∼P). *Aus* $\boldsymbol{P}\,\mathrm{I}\,\boldsymbol{g}$ *folgt* $\boldsymbol{P}^*\,\mathrm{I}\,\boldsymbol{g}^*$ *und umgekehrt.*

Beweis. Ist $\boldsymbol{g} = g(\alpha\beta)$, so ist $\boldsymbol{P}$ als $P(\alpha\beta\gamma)$ darstellbar (46). Damit gilt $\boldsymbol{P}^* = \big(L(\alpha), L(\beta), L(\gamma)\big)$ und $\boldsymbol{g}^* = \big(L(\alpha), L(\beta)\big)$ also $\boldsymbol{P}^*\,\mathrm{I}\,\boldsymbol{g}^*$.

(*) Es sei $\boldsymbol{P} = P(\alpha\beta\gamma)$ also $\boldsymbol{P}^* = \big(L(\alpha), L(\beta), L(\gamma)\big)$. Aus $\boldsymbol{P}^*\,\mathrm{I}\,\boldsymbol{g}^*$ folgt, daß $\boldsymbol{g}^*$ die Idealgeraden $\big(L(\alpha), L(\beta)\big)$, $\big(L(\beta), L(\gamma)\big)$, $\big(L(\gamma), L(\alpha)\big)$ schneidet. Damit schneidet $\boldsymbol{g}^{**} = \boldsymbol{g}$ die Polaren $g(\alpha\beta), g(\beta\gamma), g(\gamma\alpha)$ (62), inzidiert also mit $\boldsymbol{P} = P(\alpha\beta\gamma)$.

Satz 64 (A∼R) (A∼P). *Aus* $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}\,\mathrm{I}\,\boldsymbol{\epsilon}$ *folgt:* $\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{C}^*, \boldsymbol{D}^*$ *schneiden sich in einem Idealpunkt* $\boldsymbol{P}$.

(*) *Aus* $\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{C}^*, \boldsymbol{D}^*\,\mathrm{I}\,\boldsymbol{P}$ *folgt* $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}\,\mathrm{I}\,\boldsymbol{P}^*$.

Beweis. Wir können annehmen, daß keine drei der Idealpunkte $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ mit einer Idealgeraden inzidieren (63). Die Voraussetzung besagt, daß sich $(\boldsymbol{A}, \boldsymbol{B})$ und $(\boldsymbol{C}, \boldsymbol{D})$ schneiden. Nach (62) treffen sich dann auch $(\boldsymbol{A}^*, \boldsymbol{B}^*) = (\boldsymbol{A}, \boldsymbol{B})^*$ (63) und $(\boldsymbol{C}^*, \boldsymbol{D}^*) = (\boldsymbol{C}, \boldsymbol{D})^*$ in einem Idealpunkt $\boldsymbol{P}$.

(*) Wir können annehmen, daß keine drei der Idealebenen $\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{C}^*, \boldsymbol{D}^*$ mit einer Idealgeraden inzidieren (63). Aus $(\boldsymbol{A}^*, \boldsymbol{B}^*), (\boldsymbol{C}^*, \boldsymbol{D}^*)\,\mathrm{I}\,\boldsymbol{P}$ folgt nach (63) $(\boldsymbol{A}^*, \boldsymbol{B}^*)^*, (\boldsymbol{C}^*, \boldsymbol{D}^*)^*\,\mathrm{I}\,\boldsymbol{P}^*$. Nun ist wiederum nach (63) $(\boldsymbol{A}, \boldsymbol{B})^* = (\boldsymbol{A}^*, \boldsymbol{B}^*)$ also gilt $(\boldsymbol{A}^*, \boldsymbol{B}^*)^* = (\boldsymbol{A}, \boldsymbol{B})^{**} = (\boldsymbol{A}, \boldsymbol{B})$ und ebenso $(\boldsymbol{C}^*, \boldsymbol{D}^*)^* = (\boldsymbol{C}, \boldsymbol{D})$, d.h. es ist $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}\,\mathrm{I}\,\boldsymbol{P}^*$.

Wir können nun die Abbildung $\boldsymbol{\Pi}$ auch für Idealebenen erklären: Ist $\boldsymbol{\epsilon} = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$, so sei der „Pol" $\boldsymbol{\epsilon}^{\boldsymbol{\Pi}} = \boldsymbol{\epsilon}^*$ von $\boldsymbol{\epsilon}$ der Idealpunkt $(\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{C}^*)$ (nach (63) inzidieren $\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{C}^*$ nicht mit einer Idealgeraden). Satz 64 lehrt, daß die Definition von $\boldsymbol{\epsilon}^*$ nicht von der Darstellung von $\boldsymbol{\epsilon}$ (durch $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$) abhängt. Damit bildet $\boldsymbol{\Pi}$ sämtliche Idealebenen eindeutig in die Menge der Idealpunkte ab.

Satz 65 (A∼R) (A∼P). $\varepsilon^* = L(\varepsilon)$. (*) $\big(L(\varepsilon)\big)^* = \varepsilon$.

Beweis. Die eigentliche Idealebene ε sei dargestellt durch drei nichtkollineare eigentliche Idealpunkte: $\varepsilon = (A, B, C)$. Ist $A = P(\varepsilon\alpha\alpha')$, $B = P(\varepsilon\beta\beta')$, $C = P(\varepsilon\gamma\gamma')$, so folgt $L(\varepsilon)\,\mathrm{I}\,A^*, B^*, C^*$ und daraus $\varepsilon^* = L(\varepsilon)$.

(*) $L(\varepsilon)$ ist ein $P(\alpha\beta\gamma)$ mit $\alpha, \beta, \gamma \perp \varepsilon$. Es folgt $\big(L(\varepsilon)\big)^* = \big(L(\alpha), L(\beta), L(\gamma)\big)$, und aus $\varepsilon\,\mathrm{I}\,L(\alpha), L(\beta), L(\gamma)$ ergibt sich $\big(L(\varepsilon)\big)^* = \varepsilon$.

$\big(P(\alpha\beta\gamma)\big)^{**}$ ist gleich $\big(L(\alpha), L(\beta), L(\gamma)\big)^*$ also nach (65) gleich $(\alpha, \beta, \gamma) = P(\alpha\beta\gamma)$. $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})^{**}$ ist gleich $(\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{C}^*)^*$, und mit $\boldsymbol{P}^{**} = \boldsymbol{P}$ folgt $(\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{C}^*)^* = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$. Damit bildet $\boldsymbol{\Pi}$ die Mengen der Idealpunkte und der Idealebenen umkehrbar eindeutig und involutorisch aufeinander ab.

Gilt $\boldsymbol{P}\,\mathrm{I}\,\boldsymbol{\epsilon}$, so ist $\boldsymbol{\epsilon}$ als ein $(\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{R})$ darstellbar. Es folgt $\boldsymbol{\epsilon}^* = (\boldsymbol{P}^*, \boldsymbol{Q}^*, \boldsymbol{R}^*)$ und damit $\boldsymbol{\epsilon}^*\,\mathrm{I}\,\boldsymbol{P}^*$.

Gilt $\boldsymbol{g}\,\mathrm{I}\,\boldsymbol{\epsilon}$, so folgt $\boldsymbol{g}^{**}\,\mathrm{I}\,\boldsymbol{\epsilon}^{**}$, und (63*) ergibt $\boldsymbol{g}^*\,\mathrm{I}\,\boldsymbol{\epsilon}^*$.

Die Abbildung $\mathbf{\Pi}$ ist damit inzidenztreu, d.h. $\mathbf{\Pi}$ ist eine involutorische Korrelation des Idealraumes. Im Falle des Axioms P wollen wir unter dem Idealraum den metrischen Raum selbst (33) und unter der Abbildung $\mathbf{\Pi}$ den Übergang zu polaren Elementen des metrischen Raumes verstehen. Mit Satz (32) ergibt sich dann, daß unter Voraussetzung von Axiom $\sim$R in allen Fällen eine Polarität $\mathbf{\Pi}$ im Idealraum existiert, welche das Senkrechtstehen im metrischen Raum bestimmt.

Unter den *Höhen* eines Dreiecks $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ (mit den Seiten $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$) in einer Idealebene $\mathfrak{e}$ verstehen wir die idealen Verbindungsgeraden von $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ mit den idealen Schnittpunkten von $\mathfrak{e}$ mit $\boldsymbol{a}^*, \boldsymbol{b}^*, \boldsymbol{c}^*$. Aus der Gültigkeit des Höhensatzes im Bündelmodell um einen Punkt P folgt die Gültigkeit des Höhensatzes in P^*. Eine zu der Betrachtung am Ende von Nr. 2 analoge Schlußweise nach [5] ergibt dann, daß ein Büschel von Idealebenen $\boldsymbol{\xi}$ mit einem Träger, welcher mit P inzidiert, durch $\mathbf{\Pi}$ projektiv auf die Punktreihe der in P^* liegenden Pole $\boldsymbol{\xi}^*$ der Idealebenen $\boldsymbol{\xi}$ abgebildet wird. Eine Polarität, die ein Büschel projektiv abbildet, ist aber überhaupt projektiv, und wir erhalten den

Satz 66 (A$\sim$R). *Unter Voraussetzung von Axiom $\sim R$ ist der Idealraum des zu $\mathfrak{G}$ gehörigen metrischen Raumes ein ordinärer projektiv-metrischer Raum.*

4. Ergebnis der Begründung. Die Resultate von Nr. 2 und Nr. 3 besagen zusammen: Der Idealraum einer durch das Axiomensystem in § 1, 1 gekennzeichneten Bewegungsgruppe $\mathfrak{G}$ ist ein projektiv-metrischer Raum, dessen absolute Polarität $\mathbf{\Gamma}$ bzw. $\mathbf{\Pi}$ durch die Orthogonalität des metrischen Raumes bestimmt ist.

Die Spiegelung Σ_ε an einer Ebene ε des metrischen Raumes läßt jeden mit ε inzidierenden Punkt und jeden mit ε inzidierenden Idealpunkt fest (Aus $\alpha' \varepsilon \alpha = \varepsilon$ d.h. $[\alpha' \varepsilon \alpha]$ folgt $P(\varepsilon\alpha\beta) = P(\varepsilon\alpha'\beta')$) und induziert im Idealraum eine involutorische Kollineation $\mathbf{\Sigma}_\varepsilon$, welche (etwa nach [2, III, 2 Lemma 2]) projektiv ist. Da jedes σ mit $\sigma \perp \varepsilon$ bei Σ_ε ebenfalls fest bleibt, besitzt $\mathbf{\Sigma}_\varepsilon$ im Idealraum den Fixpunkt $L(\varepsilon)$, d.h. im Falle (A$\sim$R) den Pol von ε und im Falle (AR) den durch $\mathbf{\Gamma}$ der idealen Schnittgeraden von ε und $\boldsymbol{\omega}$ zugeordneten Punkt. $\mathbf{\Sigma}_\varepsilon$ ist damit eine harmonische Homologie, die mit der absoluten Polarität $\mathbf{\Gamma}$ bzw. $\mathbf{\Pi}$ vertauschbar ist. Die von den Spiegelungen an den Ebenen eines zu einem $\mathfrak{G}$ gehörigen metrischen Raumes erzeugte Gruppe ist nach (31) zu $\mathfrak{G}$ selbst isomorph. Damit ergibt ein Vergleich mit der Definition der Bewegungsgruppe eines projektiv-metrischen Raumes (Nr. 1) das abschließende

Begründungstheorem: *Durch das Axiomensystem in § 1, 1 gekennzeichnete Bewegungsgruppen metrischer Räume sind als Untergruppen von Bewegungsgruppen projektiv-metrischer Räume darstellbar.*

Mit diesem Theorem ist der Anschluß an die Cayley-Kleinsche Auffassung der Geometrie gewonnen, und die durch das Axiomensystem gegebenen Bewegungsgruppen sowie die durch sie definierten metrischen Räume sind einer Behandlung mit algebraischen Methoden zugänglich gemacht. Für den

ordinären Fall ergibt sich (etwa nach [2]), daß die Gruppen $\mathfrak{G}$ isomorph zu Untergruppen projektiv-orthogonaler Gruppen $PO_4(K, f)$ sind. Der Körper K ist hierbei kommutativ (52), die Form f eine symmetrische Bilinearform, denn da nicht jede Idealebene mit ihrem Pol inzidiert (es ist stets $\varepsilon \nexists L(\varepsilon)$), ist die Polarität $\mathbf{\Pi}$ kein Nullsystem.

Literatur

[1] Bachmann, F.: Aufbau der Geometrie aus dem Spiegelungsbegriff. Die Grund lehren der math. Wiss. **96**. Berlin-Göttingen-Heidelberg (im Druck). — [2] Baer, R.: Linear Algebra und Projective Geometry. New York 1952. — [3] Hilbert, D.: Grundlagen der Geometrie, 8. Aufl. Stuttgart 1956. — [4] Hjelmslev, J.: Einleitung in die allgemeine Kongruenzlehre. 4. Mitteilung. Det Kgl. Danske Vidensk. Selskab, Matem.-Fys. Meddelelser **22**, Nr. 6, 1—40 (1945). — [5] Lenz, H.: Zur Begründung der analytischen Geometrie. Bayr. Akad. d. Wiss., Math.-naturw. Klasse **1954**, 17—72. — [6] Schur, F.: Grundlagen der Geometrie. Berlin 1909. — [7] Thomsen, G.: Grundlagen der Elementargeometrie in gruppenalgebraischer Behandlung. Hamb. Math. Einzelschriften **15**, 1—88 (1933). — [8] Veblen, O., and J. W. Young: Projective Geometry. Boston 1910—1916. — [9] Winternitz, A.: Zur Begründung der projektiven Geometrie. Ann. of Math. **41**, 365—390 (1940).

Kiel, Mathematisches Seminar der Universität

(Eingegangen am 11. August 1958)

OSTROM, T. G., and A. WAGNER
Math. Zeitschr., 71, 186–199 (1959)

On projective and affine planes with transitive collineation groups

By
T. G. OSTROM and A. WAGNER

§ 1. Introduction

In previous papers [*12*, *13*] one of the authors has examined doubly transitive finite projective and affine planes. The main results may be summarized by the following two theorems:

THEOREM A. *Let $\mathfrak{P}$ be a finite projective plane of order n admitting a collineation group doubly transitive on the points of $\mathfrak{P}$. Then, if n is odd and not a square, $\mathfrak{P}$ is desarguesian.*

THEOREM B. *Let $\mathfrak{P}^*$ be a finite affine plane of order n admitting a collineation group doubly transitive on the affine points of $\mathfrak{P}^*$. Then, if n is odd and not a square, $\mathfrak{P}^*$ is a translation plane.*

The former result has been generalized by M. HALL and D. R. HUGHES, HUGHES [*10*], to include the case where n is even and not a square.

The main purpose of this paper is to prove these theorems without restriction, other than finiteness, on the order of the plane. We prove these results directly for all values of n, since there is little simplification in treating only those cases previously intractable.

The affine case is treated first in view of its greater simplicity. This is done in § 3.

In § 4, we consider affine planes with a collineation group transitive on all quadrangles where two of the vertices lie on the special line. This may be considered an affine analogue of transitivity on quadrangles in a projective plane, treated in WAGNER [*18*]. It is shown in Theorem 4 that such a finite affine plane is desarguesian.

In § 5, we prove our main result that all finite doubly transitive projective planes are desarguesian.

Finally in § 6, we examine finite affine planes with a collineation group transitive on affine lines. We conjecture that such planes are translation planes. If this conjecture is correct, it would imply Theorems 1 and 5 as simple corollaries. However, we have been able to prove this conjecture only by placing some restrictions on the order of the plane.

§ 2. Results required later

For the main definitions and concepts of projective planes we refer to those given in WAGNER [*19*]. In the main, these follow the usual conventions (see for example PICKERT [*15*]).

We shall use the following notation. If Π is a collineation group of a plane $\mathfrak{P}$ and $A, B, \ldots$ are points of $\mathfrak{P}$ and $l, m, \ldots$ are lines of $\mathfrak{P}$, we denote the subgroup of Π consisting of those elements of Π leaving fixed $A, B, \ldots$ and $l, m, \ldots$ by $\Pi_{A,B,\ldots,l,m,\ldots}$. The order of Π we denote by $o(\Pi)$.

Without further comment we shall make repeated use of the following well-known results. If σ is a perspectivity with centre C and axis l and ϱ is any collineation, then $\varrho^{-1}\sigma\varrho$ is a perspectivity with centre $C\varrho$ and axis $l\varrho$. If a plane is (C, l)-transitive and ϱ is any collineation, then the plane is also $(C\varrho, l\varrho)$-transitive. If Π is a permutation group on the elements of a set $\mathfrak{S}$ and Π splits $\mathfrak{S}$ into transitive classes $\mathfrak{T}_i$ containing respectively t_i objects, and if, further, T_i is an element of $\mathfrak{T}_i$, then $o(\Pi) = t_i \cdot o(\Pi_{T_i})$.

We shall also make use of certain simple arguments and a number of recent results concerning collineations. For convenience we list these.

I. Let Σ be a permutation group and Π a subgroup of Σ. Further, let $\mathfrak{F}$ be the set of objects left fixed by every element of Π. Then the normaliser of Π in Σ will permute the objects of $\mathfrak{F}$ amongst themselves.

II. (See BAER [5], Theorem 1.) Let $\mathfrak{P}$ be a finite projective plane and let ϱ be any collineation of $\mathfrak{P}$. Then the number of points left fixed by ϱ is equal to the number of lines left fixed by ϱ.

III. (See DEMBOWSKI [7], HUGHES [11], PARKER [14].) Let $\mathfrak{P}$ be a finite projective plane and let Σ be any collineation group of $\mathfrak{P}$. Then the number of transitive classes into which Σ splits the points of $\mathfrak{P}$ is equal to the number of transitive classes into which Σ splits the lines of $\mathfrak{P}$.

IV. (See BAER [4], Theorems 2 and 5.) Let $\mathfrak{P}$ be a projective plane and let σ be an involutory collineation of $\mathfrak{P}$. Then either σ is a perspectivity or σ leaves fixed pointwise a proper subplane $\mathfrak{Q}$ of $\mathfrak{P}$. If $\mathfrak{P}$ is finite and of order n, then the second alternative implies that n is a square and the order of $\mathfrak{Q}$ is $\sqrt{n}$.

V. Let $\mathfrak{P}$ be a finite projective plane of order n and let σ be a perspectivity of $\mathfrak{P}$ of order s. Then:

(1) σ is a homology if and only if $s \mid n-1$,

(2) σ is an elation if and only if $s \mid n$.

VI. (See OSTROM [12], Lemma 5.) Let $\mathfrak{P}$ be a projective plane (not necessarily finite) and let σ_1 and σ_2 be two involutory homologies of $\mathfrak{P}$ with centres C_1 and C_2 respectively and axes l_1 and l_2 respectively. If $C_1 \in l_2$ and $C_2 \in l_1$ then $\sigma_1\sigma_2$ is an involutory homology with centre $l_1 \cap l_2$ and axis C_1C_2.

VII. (See GLEASON [8], Lemma 1.6.) Let $\mathfrak{P}^*$ be a finite affine plane with special line l_∞. Further let Π be a collineation group of $\mathfrak{P}^*$ transitive on the points of l_∞. Then, if Π contains a non-trivial translation (*i.e.* an elation with axis l_∞), $\mathfrak{P}^*$ is a translation plane and Π contains the translation group of $\mathfrak{P}^*$ as a subgroup.

VIII. (See ANDRÉ [1], Theorem 4.) Let $\mathfrak{P}^*$ be a finite affine plane with special line l_∞ and let Π be a collineation group of $\mathfrak{P}^*$. Then, if for every affine

point C of $\mathfrak{P}^*$ there exists in Π a non-trivial homology with C as centre (the homology will necessarily have l_∞ as axis), $\mathfrak{P}^*$ is a translation plane and Π contains the translation group of $\mathfrak{P}^*$ as a subgroup.

§ 3. Doubly transitive finite affine planes

THEOREM 1. *Let $\mathfrak{P}^*$ be a finite affine plane and let Π be a collineation group of $\mathfrak{P}^*$ doubly transitive on the affine points of $\mathfrak{P}^*$. Then $\mathfrak{P}^*$ is a translation plane and Π contains the translation group of $\mathfrak{P}^*$ as a proper subgroup.*

We shall in the usual manner denote the special line of $\mathfrak{P}^*$ by l_∞. The order of the plane we denote by n. Let P and Q be any two affine points of $\mathfrak{P}^*$ and let $o(\Pi_{P,Q}) = k$. Since Π is doubly transitive on affine points, k does not depend on P and Q. Further, the total number of affine points is n^2, hence $o(\Pi) = n^2(n^2-1)\,k$.

We must now treat separately the case where n is even and the case where n is odd.

Case (i), n is even

Let[1]) $2^u \| n$

$2^v \| k$

$u > 0$. We now consider a Sylow 2-subgroup Λ of Π. $o(\Lambda) = 2^{2u+v}$ and Λ contains a non-trivial centre. Let λ be an element of order 2 in the centre of Λ and denote by $\mathfrak{S}$ the set of all affine points of $\mathfrak{P}^*$ left fixed by λ. Further, let s denote the number of points in $\mathfrak{S}$. It follows from result IV that λ is either a perspectivity or leaves a proper subplane of $\mathfrak{P}^*$ pointwise fixed. Since n is even, it follows from result V that if λ is a perspectivity, λ is an elation. Consequently λ is of one of the following types:

(a) λ is a translation. In this case $s = 0$.

(b) λ is an elation with an affine line as axis. In this case $\mathfrak{S}$ consists of the affine points on the axis and $s = n$.

(c) λ leaves the proper subplane $\mathfrak{Q}^*$ pointwise fixed. From result IV it follows that n is a square and the order of $\mathfrak{Q}^*$ is $\sqrt{n}$. $\mathfrak{S}$ consists of the points of $\mathfrak{Q}^*$; hence $s = n$.

We now assume that $s = n$.

From result I we see that Λ is a permutation group on the points of $\mathfrak{S}$. Hence Λ splits the points of $\mathfrak{S}$ into disjoint transitive classes. Let $\mathfrak{J}$ be one of these and let T be any point in $\mathfrak{J}$. Further, let t denote the number of points in $\mathfrak{J}$. Λ_T is a subgroup of Π_T. $o(\Pi_T) = (n^2-1)\,k$. Since $2^v \| (n^2-1)\,k$, it follows that $o(\Lambda_T) \mid 2^v$. However $o(\Lambda_T) \cdot t = o(\Lambda) = 2^{2u+v}$; hence $2^{2u} \mid t$. Thus we have shown that the number of points in every transitive class of $\mathfrak{S}$ is divisible by 2^{2u}. Therefore the number of points in $\mathfrak{S}$ is divisible by 2^{2u}, that is $2^{2u} \mid n$. But $2^u \| n$ and $u > 0$. This contradiction shows that λ is of type (a).

We have shown that Π contains a non-trivial translation. Since Π is doubly transitive on the affine points of $\mathfrak{P}^*$, it follows at once by considering

[1]) $p^u \| q$ denotes $p^u \mid q$ but $p^{u+1} \nmid q$.

conjugates of the given non-trivial translation that the group of translations contained in Π is transitive on the affine points. This completes the proof of the theorem.

Case (ii), n is odd. For this case an essentially similar argument is used as for case (i). Consequently we somewhat condense the proof.

Let $2^{u_1} || n+1$

$2^{u_2} || n-1$

$2^v || k$

$u_1, u_2 > 0$. Let Λ be a Sylow 2-subgroup of Π. $o(\Lambda) = 2^{u_1+u_2+v}$. Let λ be an element of order 2 in the centre of Λ and let $\mathfrak{S}$ denote the set of all affine points fixed by λ. s denotes the number of points in $\mathfrak{S}$.

Since n is odd, any involutory perspectivity of $\mathfrak{P}^*$ is a homology. It is easily seen that λ is of one of the following types:

(a) λ is a homology with l_∞ as axis. In this case $\mathfrak{S}$ consists only of the centre of λ and $s=1$.

(b) λ is a homology with an affine line as axis and centre on l_∞. $\mathfrak{S}$ consists of the affine points on the axis of λ and $s=n$.

(c) λ leaves a proper subplane $\mathfrak{Q}^*$ pointwise fixed. In this case n is a square, $\mathfrak{S}$ consists of the affine points of $\mathfrak{Q}^*$ and $s=n$.

We now assume that $s=n$.

Λ is again a permutation group on the points of $\mathfrak{S}$. Λ splits $\mathfrak{S}$ into disjoint transitive classes. The number of points in each class is obviously a power of 2. Since n is odd, at least one point is fixed by all elements of Λ. Denote by F one of these points. Denote by $\mathfrak{Z}$ any other transitive class of points in $\mathfrak{S}$ under Λ. Let T be any point in $\mathfrak{Z}$ and let t denote the number of points in $\mathfrak{Z}$. Then $\Lambda_T = \Lambda_{F,T}$ and $\Lambda_{F,T}$ is a subgroup of $\Pi_{F,T}$. $o(\Pi_{F,T}) = k$. Since $2^v || k$, clearly $o(\Lambda_T) | 2^v$. However, $o(\Lambda_T) \cdot t = o(\Lambda) = 2^{u_1+u_2+v}$, hence $2^{u_1+u_2} | t$. Thus the number of points in every transitive class of $\mathfrak{S}$, except the class consisting only of the point F, is divisible by $2^{u_1+u_2}$. Therefore the number of points in the set $\mathfrak{S} - F$ is divisible by $2^{u_1+u_2}$, that is $2^{u_1+u_2} | n-1$. But $2^{u_2} || n-1$ and $u_1 > 0$. This contradiction shows that λ is of type (a).

We have thus shown that there exists an non-trivial homology with axis l_∞ and centre an affine point. Since Π is certainly transitive on the affine points of $\mathfrak{P}^*$, it follows immediately by considering conjugates of λ that for every affine point of $\mathfrak{P}^*$ there exists a non-trivial homology in Π with that point as centre and l_∞ as axis. Result VIII now completes the proof.

§ 4. Planes affinely transitive on quadrangles

Let $\mathfrak{P}^*$ be an affine plane with special line l_∞. Let U, V, U', V' be any 4 points of l_∞ and let O, E, O', E' be any 4 affine points of $\mathfrak{P}^*$ such that U, V, O, E and U', V', O', E' each form a non-degenerate quadrangle[2]). Then, if there exists a collineation π of $\mathfrak{P}^*$ such that $U\pi = U'$, $V\pi = V'$, $O\pi = O'$, $E\pi = E'$, $\mathfrak{P}^*$ will be said to be *affinely transitive on quadrangles.*

[2]) A non-degenerate quadrangle is a set of 4 points such that no 3 are collinear.

Planes affinely transitive on quadrangles are of some interest for the following reason. If an affine plane $\mathfrak{P}^*$ is coordinatised by a ternary ring, a non-degenerate quadrangle U, V, O, E where $U, V \in l_\infty$ and O, E are affine points, is forced into a special rôle; see for example Pickert [15] p. 34. Clearly $\mathfrak{P}^*$ is affinely transitive on quadrangles if and only if $\mathfrak{P}^*$ may be coordinatised by isomorphic ternary rings for all choices of U, V, O, E.

We shall not here explicitly discuss the nature of a collineation group affinely transitive on quadrangles. It is intended to defer this for a subsequent paper. In our next two theorems we do not assume the planes to be finite.

Theorem 2. *Let $\mathfrak{P}^*$ be a plane affinely transitive on quadrangles. Then, if there exists a non-trivial elation of $\mathfrak{P}^*$ with an affine line as axis, $\mathfrak{P}^*$ is alternative.*

Let σ be the given elation of $\mathfrak{P}^*$ and let the affine line m be the axis of σ. Further, let A and B be any two points on l_∞. Since $\mathfrak{P}^*$ is affinely transitive on quadrangles, there exists a collineation π of $\mathfrak{P}^*$ such that $m\pi = m$, $A\pi = A$, $A\sigma\pi = B$. Then $\pi^{-1}\sigma\pi$ is an elation with axis m and $A\pi^{-1}\sigma\pi = B$. Consequently $\mathfrak{P}^*$ is $(m \cap l_\infty, m)$-transitive. Since the collineation group of $\mathfrak{P}^*$ is certainly transitive on affine lines, it follows at once by considering conjugates of the given elation that for every affine line l of $\mathfrak{P}^*$ the plane is $(l \cap l_\infty, l)$-transitive. This implies that $\mathfrak{P}^*$ may be coordinatised by a right alternative field[3]) (see for example Pickert [15] pp. 106–107, Theorem 50). It was shown by Skornyakov [17] that left alternative fields of characteristic not 2 are alternative[4]), and it was shown by San Soucie [16] that left alternative fields of characteristic 2 which coordinatise a plane with the above property are alternative. This completes the proof of Theorem 2.

Theorem 3. *Let $\mathfrak{P}^*$ be a plane affinely transitive on quadrangles. Then, if there exists an involutory homology of $\mathfrak{P}^*$ with an affine line as axis, $\mathfrak{P}^*$ is alternative.*

Let ϱ be the given homology and let m be the axis and C the centre of ϱ. Clearly $C \in l_\infty$. Let φ be a collineation of $\mathfrak{P}^*$ which fixes m but not C. Then $\varphi^{-1}\varrho\varphi$ is an involutory homology with axis m and centre $C\varphi \neq C$. It has been shown by Baer [3] that the product of two involutory homologies with common axis but distinct centres is an elation with the same axis. Hence $\varrho\varphi^{-1}\varrho\varphi$ is an elation with axis m. We may now apply Theorem 2 to complete the proof.

Theorem 4. *Let $\mathfrak{P}^*$ be a finite plane affinely transitive on quadrangles. Then $\mathfrak{P}^*$ is desarguesian.*

---

[3]) Whether the coordinatising system is a right or left alternative field depends on whether in the expression denoting incidence the coordinates of a point are written on the left or on the right. We have here implicitly adopted the same convention as that used in Wagner [19]. This is the opposite of that used in Pickert [15].

[4]) Another proof of this theorem is given in Kleinfeld: Right alternative rings. Proc. Amer. Math. Soc. 4, 939–944 (1953).

Let Σ denote the collineation group of $\mathfrak{P}^*$. We must now treat separately the case where n, the order of $\mathfrak{P}^*$, is even and the case where n is odd.

Case (i), n is even. Let A and B be any two affine points of $\mathfrak{P}^*$. We claim that if λ is an element of order 2 in the centre of a Sylow 2-subgroup of $\Sigma_{A,B}$, then λ is an elation. The proof for this assertion is contained in the proof of Theorem 5 of WAGNER [*18*] and is for that reason omitted. It is, moreover, very similar to part of the proof of Theorem 1 of this paper. λ clearly has the line AB as axis; hence using Theorem 2 the result follows.

Case (ii), n is odd. Let A and B be any two affine points of $\mathfrak{P}^*$ and let C be a point on l_∞ such that A, B, C are not collinear. Then, if λ is an element of order 2 in the centre of a Sylow 2-subgroup of $\Sigma_{A,B,C}$, λ is an involutory homology. The proof of this statement is omitted for the reason given in case (i). Clearly AB is the axis of λ; hence using Theorem 3 the result follows.

§ 5. Doubly transitive finite projective planes

In this section we shall prove[5]).

THEOREM 5. *Let $\mathfrak{P}$ be a finite projective plane and let Π be a collineation group of $\mathfrak{P}$ doubly transitive on the points of $\mathfrak{P}$. Then $\mathfrak{P}$ is desarguesian and Π contains the little projective group of $\mathfrak{P}$.*

We find it convenient to introduce the following definition: if A and B are two distinct points and l and m two distinct lines of a projective plane such that $A, B \in m$, $A \in l$, then these points and lines will be said to constitute a *figure* $\boldsymbol{C}$. Clearly the dual of a figure $\boldsymbol{C}$ is again a figure $\boldsymbol{C}$. Since $m = AB$, the above figure is already determined by A, B, l. We shall denote this figure by (A, B, l) – written in that order.

We now require a number of lemmas.

LEMMA 1. *Let $\mathfrak{P}$ be a finite projective plane. If for every figure $\boldsymbol{C}$ of $\mathfrak{P}$ there exists an involutory perspectivity fixing the elements of that figure, then $\mathfrak{P}$ is desarguesian.*

It will be necessary to treat separately the case where n, the order of $\mathfrak{P}$, is even and the case where n is odd.

Case (i), n is even. From result V it is clear that in this case every involutory perspectivity is an elation. Let (A, B, l) be any figure $\boldsymbol{C}$ and let σ be an involutory elation fixing A, B, l. Since A is contained by two distinct lines fixed by σ, namely l and AB, A is clearly the centre of σ. The dual argument shows that AB is the axis of σ. Since A and B may be any two points of $\mathfrak{P}$, it follows from a result of GLEASON [*8*] – referred to as Theorem A in WAGNER [*19*] – that $\mathfrak{P}$ is desarguesian.

Case (ii), n is odd. In this case every involutory perspectivity of $\mathfrak{P}$ is a homology. We shall first show that every line of $\mathfrak{P}$ is an axis of an involutory homology.

---

[5]) The authors are indebted to Dr. PETER DEMBOWSKI for suggesting a number of aesthetic improvements in the proof of this Theorem.

Let m be any line of $\mathfrak{P}$ and let (A, B, l) be any figure $\boldsymbol{C}$ with $A, B \in m$. Further, let σ be an involutory homology fixing A, B, l. If σ has m as axis, there is no more to prove. Consequently we now assume that σ does not have m as axis. Let C be any point on m distinct from A and B and let h be any line through C, with $h \neq m$. Since $C\sigma \neq C$, $(C, C\sigma, h)$ form a figure $\boldsymbol{C}$. Let ϱ be an involutory homology fixing $C, C\sigma, h$. If ϱ has m as axis, there is no more to prove. If m is not the axis of ϱ then:

either (a) C is the centre of ϱ and the axis of ϱ passes through $C\sigma$.
or (b) $C\sigma$ is the centre of ϱ and the axis of ϱ passes through C.

If case (a) holds, then for $\sigma^{-1}\varrho\sigma$ case (b) holds. Conversely, if case (b) holds, then for $\sigma^{-1}\varrho\sigma$ case (a) holds. From result VI it follows that $\varrho\sigma^{-1}\varrho\sigma$ is an involutory homology with axis m.

A dual argument shows that every point of $\mathfrak{P}$ is the centre of an involutory homology. It is proved in WAGNER [*19*], Theorem 2, that if in a finite projective plane every point is the centre of a non-trivial homology and every line is the axis of a non-trivial homology, then the plane is a translation plane, or the dual of such a plane. Using this result, it now follows that $\mathfrak{P}$ is a translation plane or the dual of such a plane. For our proof of Theorem 5 this is sufficient. For the sake of completeness we shall, however, deduce the further property that $\mathfrak{P}$ is desarguesian.

Let us assume that $\mathfrak{P}$ is a translation plane with special line l_∞. If the collineation group of $\mathfrak{P}$ does not fix l_∞, then $\mathfrak{P}$ may be coordinatised by a left alternative field (see PICKERT [*15*] p. 207 and p. 106, Theorem 50). As was observed in the proof of Theorem 2, this implies that $\mathfrak{P}$ is alternative. However, finite alternative planes are desarguesian (see for example PICKERT [*15*] p. 301). We now assume that every collineation of $\mathfrak{P}$ fixes l_∞. Let l be any line of $\mathfrak{P}$, with $l \neq l_\infty$, and let C be any point of $\mathfrak{P}$ with $C \in l_\infty$ but $C \notin l$. Further, let A, B be two points and m a line of $\mathfrak{P}$ such that $A, B \in l$, $A, B \notin l_\infty$, $A \in m$, $C \in m$. Clearly (A, B, m) form a figure $\boldsymbol{C}$. Let σ be an involutory homology fixing A, B, m. Since σ also fixes l_∞, it is easily seen that σ has C as centre and l as axis. Now leaving l fixed and considering all possible positions for C, it follows from a result of ANDRÉ [*1*] (referred to in WAGNER [*19*] as Theorem B corollary (i)) that $\mathfrak{P}$ is $(l \cap l_\infty, l)$-transitive. As was observed in the proof of Theorem 2, this implies that $\mathfrak{P}$ is alternative. Again, since $\mathfrak{P}$ is finite, $\mathfrak{P}$ is desarguesian.

If we assume that $\mathfrak{P}$ is the dual of a translation plane, then the dual argument to the above shows that $\mathfrak{P}$ is desarguesian. This completes the proof of Lemma 1.

LEMMA 2. *Let $\mathfrak{P}$ be a projective plane and let Π be a collineation group of $\mathfrak{P}$ doubly transitive on the points of $\mathfrak{P}$. Then Π is transitive on figures $\boldsymbol{C}$.*

Let the order of $\mathfrak{P}$ be n and let S be any point of $\mathfrak{P}$. Π_S splits the points of $\mathfrak{P}$ into two transitive classes; S is one class and all the other points of $\mathfrak{P}$ form the other class. Result III shows that Π_S splits the lines of $\mathfrak{P}$ into two transitive classes also. Since clearly the $n+1$ lines through S form one class, the n^2

lines of $\mathfrak{P}$ not passing through S constitute the other class. Hence $n+1 \mid o(\Pi_S)$ and $n^2 \mid o(\Pi_S)$. Since $(n+1, n) = 1$, it follows that $n^2(n+1) \mid o(\Pi_S)$. Let $o(\Pi_S) = n^2(n+1)\,k$. Since Π is transitive on the points of $\mathfrak{P}$, $o(\Pi_S)$ is independent of the choice of S.

Now let T be any point of $\mathfrak{P}$ distinct from S. Since T lies in a transitive class of $n(n+1)$ points under Π_S, $o(\Pi_{S,T}) = nk$. Further, let m be any line of $\mathfrak{P}$ such that $S \in m$, $m \neq ST$. (S, T, m) form a figure $\boldsymbol{C}$. We have seen above that m lies in a transitive class of n^2 elements under Π_T. Hence $o(\Pi_{T,m}) = (n+1)\,k$. But $\Pi_{S,T,m}$ is a subgroup of $\Pi_{S,T}$ and also a subgroup of $\Pi_{T,m}$, hence $o(\Pi_{S,T,m}) \mid nk$ and also $o(\Pi_{S,T,m}) \mid (n+1)\,k$. Since $(n, n+1) = 1$, $o(\Pi_{S,T,m}) \mid k$. Let m lie in a transitive class containing s elements under $\Pi_{S,T}$. Then $s \cdot o(\Pi_{S,T,m}) = o(\Pi_{S,T}) = nk$, hence $n \mid s$. Since $0 < s \leqq n$ it follows that $s = n$. Thus we have shown that $\Pi_{S,T}$ is transitive on the lines through S distinct from ST.

We may now complete the proof of the Lemma. Let (A, B, l) and (S, T, m) each form a figure $\boldsymbol{C}$. Since Π is doubly transitive on points of $\mathfrak{P}$, there exists an element $\varrho \in \Pi$ such that $A\varrho = S$, $B\varrho = T$, and clearly $S \in l\varrho$ and $l\varrho \neq ST$. Above we have shown that there exists an element $\sigma \in \Pi_{S,T}$ such that $l\varrho\sigma = m$. Hence the collineation $\varrho\sigma$ maps A onto S, B onto T and l onto m, which establishes the Lemma.

COROLLARY. *Π is doubly transitive on the lines of $\mathfrak{P}$.*

This corollary has also been proved by DEMBOWSKI [7]. The corollary shows that in defining double transitivity for a projective plane it is immaterial whether we define the transitivity as being on points or on lines.

LEMMA 3. *Let $\mathfrak{P}$ be a finite projective plane and let Π be a collineation group of $\mathfrak{P}$ doubly transitive on the points of $\mathfrak{P}$. Further let A and B be any two points and l any line of $\mathfrak{P}$ such that (A, B, l) form a figure $\boldsymbol{C}$. If the order of $\mathfrak{P}$ is n and $o(\Pi_{A,B,l}) = k$ then:*

$$
\begin{aligned}
o(\Pi) &= n^2(n^2+n+1)(n+1)\,k \\
o(\Pi_A) &= n^2(n+1)\,k \\
o(\Pi_{A,B}) &= nk \\
o(\Pi_{A,l}) &= n^2 k \\
o(\Pi_{B,l}) &= (n+1)\,k.
\end{aligned}
$$

The proof of this Lemma follows immediately from the proof of Lemma 2.

We now make the following further definitions. Let $\mathfrak{P}$ be a projective plane and let Π be a collineation group of $\mathfrak{P}$. A subplane $\mathfrak{Q}$ of $\mathfrak{P}$ will be said to be a *2-subplane of $\mathfrak{P}$ with respect to Π* if there exists a 2-subgroup of Π which leaves $\mathfrak{Q}$, and only $\mathfrak{Q}$, pointwise fixed. A 2-subplane of $\mathfrak{P}$ with respect to the group of all collineations of $\mathfrak{P}$ will be referred to simply as a *2-subplane of $\mathfrak{P}$*. We observe that since the identity is a 2-group, $\mathfrak{P}$ is always a 2-subplane of $\mathfrak{P}$. In the natural manner we shall say that $\mathfrak{Q}$ is a *minimal 2-subplane of $\mathfrak{P}$ with respect to Π* if $\mathfrak{Q}$ is a 2-subplane of $\mathfrak{P}$ with respect to Π and no proper subplane of $\mathfrak{Q}$ has this property. We note that if $\mathfrak{P}$ is finite, a minimal

2-subplane of $\mathfrak{P}$ exists for all collineation groups of $\mathfrak{P}$. This property is not shared by all infinite planes.

THEOREM 6. *Let $\mathfrak{P}$ be a finite projective plane and let $\mathfrak{Q}$ be a 2-subplane of $\mathfrak{P}$ with respect to a collineation group Π of $\mathfrak{P}$. If the order of $\mathfrak{P}$ is n and the order of $\mathfrak{Q}$ is n', then $n = (n')^{2^g}$ where g is some integer.*

Let Φ be a 2-subgroup of Π leaving $\mathfrak{Q}$, and only $\mathfrak{Q}$, pointwise fixed. If Φ is the identity, $\mathfrak{Q} = \mathfrak{P}$ and there is no more to prove. We now assume that $\Phi \neq 1$. Let φ be an element of order 2 in the centre of Φ. From result IV it follows that φ is either a perspectivity or leaves a proper subplane $\mathfrak{P}_1$ of $\mathfrak{P}$ pointwise fixed, where $\sqrt{n}$ is the order of $\mathfrak{P}_1$. Since Φ, hence φ, leaves $\mathfrak{Q}$ pointwise fixed, the former alternative is impossible. Result I shows that Φ is a permutation group on the points of $\mathfrak{P}_1$. It is easily verified that if Σ_1 is the subgroup of Φ leaving $\mathfrak{P}_1$, and only $\mathfrak{P}_1$, pointwise fixed, then Φ regarded now as a collineation group of $\mathfrak{P}_1$ is isomorphic to $\Phi_1 = \Phi/\Sigma_1$. If Φ_1 is the identity the result follows. If $\Phi_1 \neq 1$ we may repeat the preceding argument by replacing $\mathfrak{P}$ by $\mathfrak{P}_1$ and Φ by Φ_1. This process may be repeated until, say at the g-th step, $\mathfrak{Q} = \mathfrak{P}_g$. Then $n' = n^{2^{-g}}$.

LEMMA 4. *Let $\mathfrak{P}$ be a finite projective plane and let Π be a collineation group of $\mathfrak{P}$ doubly transitive on the points of $\mathfrak{P}$. Further, let $\mathfrak{Q}$ be a minimal 2-subplane of $\mathfrak{P}$ with respect to Π. Then for every figure $\boldsymbol{C}$ in $\mathfrak{Q}$ there exists an element of Π which is an involutory perspectivity of $\mathfrak{Q}$ and which fixes that figure elementwise.*

Let n and k be defined as in the enunciation of Lemma 3. Further, let

$$2^s \,||\, n^2(n^2 + n + 1)(n + 1)$$
$$2^t \,||\, k.$$

Clearly $s > 0$. Now let Φ be a maximal 2-subgroup of Π leaving $\mathfrak{Q}$, and only $\mathfrak{Q}$, pointwise fixed. That is, no 2-subgroup of Π properly containing Φ leaves $\mathfrak{Q}$, and only $\mathfrak{Q}$, pointwise fixed. Since Π is of finite order, Φ certainly exists. (In the case $\mathfrak{Q} = \mathfrak{P}$, Φ is the identity.) Let $o(\Phi) = 2^u$.

Let Λ be a Sylow 2-subgroup of Π. $o(\Lambda) = 2^{s+t}$. Since $s > 0$, Λ cannot leave a figure $\boldsymbol{C}$ elementwise fixed. However, Φ certainly leaves a figure $\boldsymbol{C}$ elementwise fixed. Hence Φ cannot be a Sylow 2-subgroup of Π. Now let Φ^* be a subgroup of Π of order 2^{u+1} and such that Φ is a subgroup of Φ^*. (Such a subgroup Φ^* can always be found; see for example BURNSIDE [6] p. 151.) Since Φ is of index 2 in Φ^*, Φ is a normal subgroup of Φ^*. Further, since Φ leaves $\mathfrak{Q}$ pointwise fixed, it follows from result I that Φ^* is a permutation group on the points of $\mathfrak{Q}$. Φ was defined to be a maximal 2-subgroup of Π leaving $\mathfrak{Q}$ pointwise fixed, consequently Φ^* cannot leave $\mathfrak{Q}$ pointwise fixed. As remarked in the proof of Theorem 6, Φ^* regarded as a collineation group of $\mathfrak{Q}$ is isomorphic to Φ^*/Φ. Φ^*/Φ is the cyclic group of order 2. Let φ^* be the element of order 2 in Φ^*/Φ. If φ^* were to leave a proper subplane of $\mathfrak{Q}$ pointwise fixed, then Φ^* would leave a proper subplane of $\mathfrak{Q}$ pointwise

fixed, which contradicts the definition of $\mathfrak{Q}$. Consequently φ^* is an involutory perspectivity of $\mathfrak{Q}$.

φ^* must leave fixed elementwise a figure C in $\mathfrak{Q}$. This may easily be seen, regardless of whether φ^* is an elation or a homology of $\mathfrak{Q}$, as follows. Let the point C be the centre of φ^*, considered as a perspectivity of $\mathfrak{Q}$, let D be a point of $\mathfrak{Q}$ on the axis of φ^*, with $C \neq D$, and let m be a line of $\mathfrak{Q}$ with $C \in m$ and $m \neq CD$. Then (C, D, m) form a figure C. Hence Φ^* leaves fixed elementwise a figure C. Consequently $o(\Phi^*) \mid k$ thus $u < t$.

Now let (S, T, l) be any figure C in $\mathfrak{Q}$. Clearly Φ is a subgroup of $\Pi_{S,T,l}$. Since $o(\Pi_{S,T,l}) = k$, Φ cannot be a Sylow 2-subgroup of $\Pi_{S,T,l}$. We now choose a 2-subgroup Φ^* of $\Pi_{S,T,l}$ such that $o(\Phi^*) = 2^{u+1}$ and Φ is a subgroup of Φ^*. We may now employ the preceding argument to show that there exists an element of order 2 in Φ^*/Φ which is an involutory perspectivity of $\mathfrak{Q}$. Since φ^* may be represented by an element of $\Pi_{S,T,l}$, φ^* leaves S, T, l fixed. This completes the proof of Lemma 4.

LEMMA 5. *Let $\mathfrak{P}$ be a finite projective plane admitting a collineation group doubly transitive on the points of $\mathfrak{P}$. Then the order of $\mathfrak{P}$ is a power of a prime.*

This Lemma follows at once by combining the results of Lemmas 1 and 4, and Theorem 6, and by observing that the order of a finite desarguesian plane is a power of a prime.

LEMMA 6. *Let $\mathfrak{P}$ be a finite projective plane and let Π be a collineation group of $\mathfrak{P}$ doubly transitive on the points of $\mathfrak{P}$. Then Π contains a non-trivial elation.*

Let n and k be defined as in the enunciation of Lemma 3. Now let l and m be any two distinct lines of $\mathfrak{P}$ and let $S = l \cap m$. Denote by $\mathfrak{L}$ the set of points of l distinct from S and by $\mathfrak{M}$ the set of points of m distinct from S. By dualising the result of Lemma 3, it is seen that $o(\Pi_{l,m}) = nk$. In view of Lemma 5, n is a power of a prime.

Let $n = p^s$ where p is a prime and $p^t \| k$.

We shall denote by Λ a Sylow p-subgroup of $\Pi_{l,m}$. $o(\Lambda) = p^{s+t}$. Λ splits the points of $\mathfrak{P}$ into transitive classes. S forms a transitive class. Now let T be any other point of $\mathfrak{P}$ distinct from S and let T lie in a transitive class containing q points. If $T \notin l$, Λ_T is a subgroup of $\Pi_{S,T,l}$, while if $T \in l$, Λ_T is a subgroup of $\Pi_{S,T,m}$. $o(\Pi_{S,T,l}) = o(\Pi_{S,T,m}) = k$, hence $o(\Lambda_T) \mid k$, thus $o(\Lambda_T) \mid p^t$. Since $q \cdot o(\Lambda_T) = o(\Lambda) = p^{s+t}$, it follows that $p^s \mid q$, that is $n \mid q$.

Consequently, all the points of $\mathfrak{P}$, except S, lie in transitive classes under Λ containing at least n points. In particular, Λ is transitive on $\mathfrak{L}$ and $\mathfrak{M}$.

Λ contains a non-trivial centre. Lst λ be an element of order p in the centre of Λ. Now λ leaves fixed at least two lines of $\mathfrak{P}$, namely l and m. From result III it follows that λ must fix at least two points of $\mathfrak{P}$. One of these is S and let the other (or one of the others) be T. One or other of the following three cases must occur.

(i) $T \in l$,
(ii) $T \in m$,
(iii) $T \notin l, m$.

We first note that in view of result I, λ fixes every point in the same transitive class under Λ as T.

Case (i). Since $\mathfrak{L}$ forms a single transitive class under Λ, λ leaves every point of $\mathfrak{L}$, hence l, fixed. Consequently λ is a perspectivity with axis l.

Case (ii). From the same argument as was used for case (i), λ is a perspectivity with axis m.

Case (iii). Let U be any other point in the same transitive class under Λ as T. If TU does not contain S, let TU meet l and m in the points V, W respectively. Since T and U are left fixed by λ, V and W are fixed also. From cases (i) and (ii), we see that this implies that λ is a perspectivity with axis l and m. Consequently λ is the identity, which contradicts the definition of λ.

Thus for every point U in the same transitive class as T under Λ, S, T, U are collinear. Since T lies in a transitive class of at least n elements, it follows that λ fixes all points of the line ST. Hence λ is a perspectivity with axis ST.

Thus we have shown that λ is a perspectivity of order p. The fact that λ is an elation follows from result V. This completes the proof of Lemma 6.

It was shown in WAGNER [*19*], Theorem 3, that if $\mathfrak{P}$ is a finite projective plane and Π is a collineation group of $\mathfrak{P}$ transitive on the points of $\mathfrak{P}$ and if, further, Π contains a non-trivial perspectivity, then $\mathfrak{P}$ is desarguesian and Π contains the little projective group of $\mathfrak{P}$. Theorem 5 now follows immediately by combining this result with Lemma 6.

We observe that the little projective group of a desarguesian plane (or, more generally, of an alternative plane) is certainly doubly transitive on the points of the plane. Consequently Π may coincide with the little projective group.

§ 6. Transitive finite affine planes

In this section we examine finite affine planes with a collineation group transitive on the affine lines. We shall show that such planes are translation planes except, possibly, when the order of the plane is the square of a composite odd number (that is, has at least two distinct prime factors). The methods we shall use are essentially the same as those employed previously, consequently we shall summarize the proofs.

LEMMA 7. *Let $\mathfrak{P}^*$ be a finite affine plane of order n and let Π be a collineation group transitive on the affine lines of $\mathfrak{P}^*$. Then*

(i) Π is transitive on the affine points of $\mathfrak{P}^*$.

(ii) $o(\Pi) = n^2(n+1)\,k$, where k is some integer.

(iii) If A is some point of l_∞, the special line of $\mathfrak{P}^*$, then Π_A is transitive on the affine points of $\mathfrak{P}^*$.

Part (i) follows directly from result III and part (ii) is an immediate consequence of part (i). To prove part (iii) we first observe that $o(\Pi_A) = n^2 k$ and $o(\Pi_S) = (n+1)\,k$, where S is any affine point of $\mathfrak{P}^*$. If s is the number

of points in the transitive class under Π_A containing S, then $o(\Pi_{A,S}) = \frac{n^2 k}{s}$. Since $\Pi_{A,S}$ is a subgroup of Π_S, $\frac{n^2 k}{s} \mid (n+1)\, k$. However, $(n, n+1) = 1$, hence $n^2 \mid s$. This implies that $s = n^2$ and completes the proof of Lemma 7.

LEMMA 8. *Let $\mathfrak{P}^*$ be a finite affine plane and let Π be a collineation group transitive on the affine points of $\mathfrak{P}^*$. If A is any point on l_∞, then Π_A is transitive on the affine lines through A.*

Let Π split the points on l_∞ into s transitive classes. Hence the total number of transitive classes of points under Π is $s+1$. From result III it follows that there are also $s+1$ transitive classes of lines. l_∞ is one of these, and it is easily seen that all the affine lines intersecting l_∞ in points of the same transitive class must themselves lie in one transitive class. In particular, Π is transitive on the affine lines through A. The Lemma now follows immediately.

THEOREM 7. *Let $\mathfrak{P}^*$ be a finite affine plane, of order n, and Π a collineation group transitive on the affine lines of $\mathfrak{P}^*$. Then if*

(i) n is even

or (ii) n is odd and not a square

or (iii) n is a power of a prime

$\mathfrak{P}^*$ is a translation plane and Π contains the translation group of $\mathfrak{P}^*$ as a proper subgroup.

Part (i). If λ is an element of order 2 in the centre of a Sylow 2-subgroup of Π, then an argument entirely similar to that of Theorem 1 shows that λ is an elation with axis l_∞. The Theorem now follows immediately by using result VII.

Part (ii). In this case all involutions are homologies. To show that there always exists a homology with axis l_∞ we use the methods of OSTROM [*12*], Lemma 8. Since Π is transitive on affine points, clearly every affine point of $\mathfrak{P}^*$ is the centre of a homology with axis l_∞. The Theorem now follows from result VIII.

Part (iii). Let $n = p^s$, $p^t \| k$

where k is defined as in Lemma 7. Further, let A_1 be any point on l_∞ and Λ a Sylow p-subgroup of Π_{A_1}. $o(\Lambda) = p^{2s+t}$. Let U be any affine point of $\mathfrak{P}^*$ and let the transitive class under Λ containing U consist of p^u points. Then $o(\Lambda_U) = p^{2s+t-u}$. It was shown in Lemma 7 that $o(\Pi_{A_1,U}) = k$. Since Λ_U is a subgroup of $\Pi_{A_1,U}$, $p^{2s+t-u} \mid k$, hence $p^u = p^{2s}$. This shows that Λ is transitive on the affine points of $\mathfrak{P}^*$. Now let A_2 be any other point on l_∞ such that Λ_{A_2} is of maximal order. In view of Lemma 8, Λ_{A_2} is transitive on the affine lines through A_2. Also Λ_{A_2} is transitive on the affine lines through A_1 since $\Lambda_{A_2} = \Lambda'_{A_1}$, where Λ' is some Sylow p-subgroup of Π_{A_1}. Now let $A_1, A_2, \ldots, A_i$ be all the points on l_∞ left fixed by Λ_{A_1}. Λ_{A_2} is a subgroup of Λ_{A_j} $(j = 2, \ldots, i)$,

hence by the maximal property of Λ_{A_1}, $\Lambda_{A_1} = \Lambda_{A_j}$. Consequently Λ_{A_1} is transitive on all the affine lines through A_j. Let λ be an element of order p in the centre of Λ_{A_1}. λ leaves fixed at least two points, hence must fix at least two lines. One of these is l_∞; denote the other (or one of the others) by l. If l intersects l_∞ at A_j, where j belongs to the set $1, 2, \ldots, i$, then λ fixes every affine line through A_j since Λ_{A_1} is transitive on the affine lines through A_j and λ is normal in Λ_{A_1}. Consequently, λ is a perspectivity with centre A_j. If l does not intersect l_∞ at a point left fixed by Λ_{A_1}, then l must lie in the same transitive class under Λ_{A_1} as a line m, where l and m intersect in an affine point. Denote this point by P. Since l and m are left fixed by λ, P is also left fixed by λ. However, this implies that A_1P and A_2P are left fixed by λ, hence by the preceding argument λ is a perspectivity with centres A_1 and A_2, that is λ is the identity. This contradiction shows that only the former alternative is possible.

We have shown that λ contains an elation with axis l_∞. Result VII now completes the proof.

§ 7. Conclusion

In view of Theorem 5, we may regard the structure of doubly transitive finite projective planes and their collineation groups as being completely determined. The same is not true of doubly transitive finite affine planes. The only example known to the authors of a non-desarguesian plane of this type is the translation plane of order 9 (see for example ANDRÉ [2], where this plane is referred to as the *exception plane*). It is not known whether this is the only non-desarguesian plane with this property. However, it is pointed out in OSTROM [12] that any other non-desarguesian plane of this type must have the property that all the coordinatising rings of the plane have non-associative multiplication and possess only one distributive law.

It is natural to attempt to remove the restrictions on the order of the plane from Theorem 7 by methods analogous to those used in the proof of Theorem 5. A minimal 2-subplane (analogous to the plane Ω of § 5) may easily be shown to possess many involutory perspectivities. However, we have not been able to show that these perspectivities are necessarily distributed in a manner which would enable us to deduce that the minimal 2-subplane is a translation plane, or at least of prime power order. By very lengthy methods we may somewhat weaken the restrictions on the order of the plane in Theorem 7, but this seems of little interest. The smallest plane for which all our methods fail has order 4, 225.

Finally, the more restricted problem of determining all finite translation planes with a collineation group transitive on affine lines would seem to be of interest. This is, of course, equivalent to determining the finite translation planes with a collineation group transitive on the points of the special line (l_∞). D. R. HUGHES has pointed out to us that he has shown that, except for the plane of order 9, no translation plane based on a Hall Veblen-Wedderburn Algebra, see HALL [9] appendix II, has this property.

References

[1] André, J.: Über Perspektivitäten in endlichen projektiven Ebenen. Arch. Math. **6**, 29–32 (1954). – [2] André, J.: Projektive Ebenen über Festkörpern. Math. Z. **62**, 137–160 (1955). – [3] Baer, R.: The fundamental theorems of elementary geometry. Trans. Amer. Math. Soc. **56**, 94–129 (1944). – [4] Baer, R.: Projectivities with fixed points on every line of the plane. Bull. Amer. Math. Soc. **52**, 273–286 (1946). – [5] Baer, R.: Projectivities of finite projective planes. Amer. J. Math. **69**, 653–684 (1947). – [6] Burnside, W.: Theory of groups of finite order. Cambridge 1911. – [7] Dembowski, H. P.: Verallgemeinerungen von Transitivitätsklassen endlicher projektiver Ebenen. Math. Z. **69**, 59–89 (1958). – [8] Gleason, A. M.: Finite Fano planes. Amer. J. Math. **78**, 797–807 (1956). – [9] Hall, M.: Projective planes. Trans. Amer. Math. Soc. **54**, 229–277 (1943). – [10] Hughes, D. R.: Generalized incidence matrices over group algebras. Illinois J. Math. **1**, 545–551 (1957). – [11] Hughes, D. R.: Collineations and generalized incidence matrices. Trans. Amer. Math. Soc. **86**, 284–296 (1957). – [12] Ostrom, T. G.: Double transitivity in finite projective planes. Canad. J. Math. **8**, 563–567 (1956). – [13] Ostrom, T. G.: Transitivities in projective planes. Canad. J. Math. **9**, 389–399 (1957). – [14] Parker, E. T.: On collineations of symmetric designs. Proc. Amer. Math. Soc. **8**, 350–351 (1957). – [15] Pickert, G.: Projektive Ebenen. Heidelberg 1955. – [16] San Soucie, R. L.: Right alternative division rings of characteristic two. Proc. Amer. Math. Soc. **6**, 291–296 (1955). – [17] Skornyakov, L. A.: Right alternative fields. [Russian.] Izv. Akad. Nauk SSSR., Ser. Mat. **15**, 177–184 (1951). – [18] Wagner, A.: On projective planes transitive on quadrangles. J. Lond. Math. Soc. **33**, 25–33 (1958). – [19] Wagner, A.: On perspectivities of finite projective planes. Math. Z. **71**, 113–123 (1959).

Montana State University, Missoula, Montana

The Ohio State University, Columbus, Ohio
Imperial College, London

(Eingegangen am 26. August 1958)

TACHIKAWA, H.
Math. Zeitschr. 71, 200–222 (1959)

On Rings for which Every Indecomposable Right Module has a Unique Maximal Submodule

By
HIROYUKI TACHIKAWA

Introduction. Let A be a ring possessing a unit element 1 and satisfying the minimum condition on right (resp. left) ideals. A right (resp. left) A-module M will be said to be quasi-primitive, if M has a unique maximal submodule, i.e. M is isomorphic to a homomorphic image of a primitive right (resp. left) ideal. If every finitely generated indecomposable right (resp. left) A-module is quasi-primitive, we shall call A a ring of *strong right* (resp. *lefft*) *cyclic representation type*, abbreviated by ring of SRCRT (resp. SLCRT) or SRCRT (resp. SLCRT) ring[1]. In the case of commutative rings the notions of quasi-primitive module and indecomposable cyclic module are equivalent to each other, and it has been established by KOETHE [*1*] that A is of CRT if and only if A is a principal ideal ring.

Extending KOETHE'S result to the case of non-commutative rings T. NAKAYAMA [*4*, *5*] has proved that for a ring A to be of SRCRT and SLCRT it is necessary and sufficient that A is a generalized uni-serial ring. At the same time he has shown an example which is a SRCRT ring but not a SLCRT ring. The purpose of this paper is to determine the structure of a ring of SRCRT which is not necessarily of SLCRT.

For the case of algebras our characterization of SRCRT will be stated as follows: Let A be a finite dimensional algebra over a field P, N the radical of A and $e_\varkappa, e_\lambda, e_\mu, \ldots$ are primitive idempotents of A. Then in order that A may be of SRCRT it is necessary and sufficient that the following conditions are satisfied:

I. *A is a left generalized uni-serial algebra, i.e. every primitive left A-module $Ae_\varkappa$ has a unique composition series.*

II. *Let L_1 and L_2 be quasi-primitive left A-modules with the same composition length, and M_1 and M_2 submodules of L_1 and L_2 respectively. If each of M_i, $i=1, 2$ contains the S^2-part of L_i (= the annihilator of N^2 in L_i), then every isomorphism θ of M_1 onto M_2 is always extendable to an isomorphism Θ of L_1 onto L_2.*

III. *Assume that $Ne_\varkappa/N^2e_\varkappa \approx Ne_\lambda/N^2e_\lambda \approx Ne_\mu/N^2e_\mu$ and $Ae_\lambda \not\approx Ae_\mu$, then $Ae_\varkappa$ is isomorphic to either Ae_λ or Ae_μ.*

IV. *Assume that $Ne_\lambda/N^2e_\lambda \approx Ae_\varkappa/Ne_\varkappa$, then $e_\varkappa Ne_\lambda/e_\varkappa N^2e_\lambda$ is considered as a right vector space over a quasi-field $e_\lambda Ae_\lambda/e_\lambda Ne_\lambda$ and its dimension $(e_\varkappa Ne_\lambda/e_\varkappa N^2e_\lambda$:*

[1]) The denomination is due to T. YOSHII. — The results of this paper were announced by the author at the spring meeting of Math. Soc. of Japan on 30th. May, 1958.

$e_\lambda A e_\lambda / e_\lambda N e_\lambda) \leqq 2$. *Particularly if the equality holds and* $N e_\lambda / N^2 e_\lambda \approx N e_\mu / N^2 e_\mu$, *then there results that* $A e_\lambda \approx A e_\mu$.

In order to state the characterization for rings with minimum condition we have only to replace A in the above conditions I–IV by the inverse A-homomorphism ring B of a finitely generated, fully faithful[2]), injective right A-module.

In § 1 the interlacing of modules which was introduced by SNAPPER [6] and plays an important role in this paper will be investigated. By the duality of character modules we shall reduce our problem to the investigation for rings possessing a property such that every indecomposable module has a unique minimal submodule and it will be shown in § 2 that for this ring every quasi-primitive module is uniquely determined up to monomorphism by its S^2-part. §§ 3, 4 and 5 will be devoted to the discussion of necessary and sufficient conditions for rings of SRCRT. As an application of our results we shall prove that the inverse A-homomorphism ring of a fully faithful, injective right A-module is generalized uni-serial if and only if A is generalized uni-serial.

Throughout this paper rings will be assumed to be associative rings with unit elements, and by a module with a ring as a left (resp. right) operator domain we shall always mean a finitely generated module on which the unit element of ring acts as the identity operator.

§ 1. Interlacing of modules

Let B be a ring with unit element and satisfy the minimum condition on left ideals. Let M_i, $i=1, 2$, be left B-modules, and S_i, $i=1, 2$, submodules of M_i respectively. If S_1 and S_2 are isomorphic to each other by a map $\theta: S_1 \to S_2$. Then we can construct a new left B-module from M_1 and M_2 by identifying each element x of S_1 with the corresponding element $\theta(x)$ of S_2. This new left B-module M is called the interlacing module (or only, interlacing) of M_1 and M_2 obtained by using θ as lacing isomorphism and denoted by $\mathrm{Int}_\theta(M_1, M_2)$. Then there exist injections α and β of M_1 and M_2 respectively into M such that

$$\alpha(x_1) = \beta(x_2) \quad \text{if and only if} \quad x_2 = \theta(x_1). \tag{1}$$

If L is a sum of left B-modules L_1 and L_2, L may be considered as an interlacing of L_1 and L_2 obtained using the identity mapping of $L_1 \cap L_2$ onto itself as lacing isomorphism. Hence we have

LEMMA 1.1. *Every finitely generated left B-module is isomorphic to an interlacing module of a finite number of quasi-primitive left B-modules.*

LEMMA 1.2. *A left B-module L is injective if and only if for any quasi-primitive left B-module P every interlacing module of L with P has a direct summand which is isomorphic to L.*

[2]) A right (resp. left) A-module M is said to be fully faithful, if it is faithful and every simple right (resp. left) A-module is isomorphic to a submodule of M.

Proof. Let Q be the injective hull[3]) of L. If $L \neq Q$, then there exists an element $e_\varkappa u$ of Q such that $e_\lambda u$ does not belong to L, where $e_\varkappa$ is a primitive idempotent of B. Since $Be_\varkappa u + L$ may be considered as an interlacing module of $Be_\varkappa u$ and L, it follows from the assumption that $Be_\varkappa u + L = Be_\varkappa u' \oplus L$ where $Be_\varkappa u \neq 0$. Then the socles[3]) of Q and L are distinct to each other. This contradicts to that Q is the injective hull of L.

Now let T_1 and T_2 be submodules of M_1 and M_2 such that T_1 and T_2 contains properly submodules S_1 and S_2 of M_1 and M_2 respectively. If for a homomorphism $\varphi: S_1 \to S_2$ there exists a homomorphism $\Phi: T_1 \to T_2$ which coincides with φ on S_1, then φ is said to be extendable to Φ and Φ is called an extension of φ. If φ has no extension, then φ is said to be maximal.

Let us denote by N the radical of B. For a left B-module L we shall denote by $S^i(L)$, $i = 1, 2, \ldots$ the annihilator of N^i in L and call $S^i(L)$ the S^i-part of L. Especially S^1-part of L is the socle of L.

LEMMA 1.3. *Let M_1 and M_2 be left B-modules. For S_i, T_i of M_i, $i = 1, 2$, and for isomorphisms φ, ψ we assume that the followings hold*:

(i) *S_i, $i = 1, 2$ is the S^1-part of M_i,*

(ii) *φ is a monomorphism of $S^1(M_1)$ into $S^1(M_2)$,*

(iii) *ψ is an isomorphism of T_1 onto T_2 such that ψ is an extension of φ.*

Let us denote by L the interlacing of M_1 and M_2 obtained by using ψ as the lace isomorphism. Then it holds that $S^1(L) = S_2\beta$ $(\supseteq S_1\alpha)$ if and only if ψ is maximal.

Proof. If $S^1(L) \neq S_2\beta$, then there exists a simple submodule M of L such that the intersection of M and $S_2\beta$ is a zero element. Let x be a generator of M, then we have

$$x = x_1\alpha + x_2\beta, \quad x_1 \in M_1, \ x_2 \in M_2. \tag{2}$$

Now assume that $x_1\alpha \in T_1\alpha$, then from the definition of the interlacing there exists an element $x_1' \in T_2$ such that $x_1\alpha = x_1'\beta$. Therefore $x = (x_1' + x_2)\beta$ and $M = Bx \subseteq M_2\beta$. Hence $M \cap S^1(M_2\beta) = M \cap S_2\beta \neq 0$, and this is a contradiction. Thus we obtain that

$$x_1\alpha \notin T_1\alpha \quad \text{and similarly} \quad x_2\beta \notin T_2\beta. \tag{3}$$

Now we shall define a homomorphism $\xi: Bx_1\alpha \to Bx_2\beta$ by putting

$$\xi(bx_1\alpha) = -bx_2\beta \quad \text{for all } b \text{ of } B. \tag{4}$$

We must show first that ξ is really well defined. If $bx_1\alpha = 0$ and $bx_2\beta \neq 0$ for some b of B, then $bx = bx_1\alpha + bx_2\beta = bx_2\beta \neq 0$. Since Bx is simple,

$$M = Bx = Bbx = Bbx_2\beta, \tag{5}$$

and thus M is contained in $S_2\beta$. This contradicts to the assumption that

[3]) Im previous papers [2, 3, 7] we used the term "semisimple part" in place of "socles". By injective hull of L we shall mean a minimum injective module which contains L.

$M \cap S_2\beta = 0$. Hence if $bx_1 = 0$, then $bx_2 = 0$. This shows that ξ is well defined and further ξ is an isomorphism.

Assume that $Bx_1\alpha \cap T_1\alpha = 0$ and $Bx_2\beta \cap T_2\beta = 0$, then by using the isomorphism ξ we have an extension of ψ which is an isomorphism. So assume that $Bx_1\alpha \cap T_1\alpha \neq 0$, then there exists a non-zero element $bx_1\alpha$ contained in $T_1\alpha$ and hence there exists $y \in T_2$ such that $bx_1\alpha = y\beta$, when $bx = bx_1\alpha + bx_2\beta = (y + bx_2)\beta$. Thus we again get to a contradiction unless $bx = 0$. Assume that $bx = 0$, then $\psi(bx_1\alpha) = -bx_2\beta$. Therefore ψ coincides with ξ on the domain $Bx_1\alpha \cap T_1\alpha$. For the case where $Bx_2\beta \cap T_2\beta \neq 0$ we have similarly an extension of ψ. Thus we completes the proof of "if part".

Assume that ψ is not maximal. Let $\Psi: U_1 \to U_2$ be a maximal extension of ψ. Then the submodule V of L consist of all elements written as $x\alpha - \Psi(x)\beta$ for all $x \in U_1$ is not zero-module and $M_2\beta \cap V = 0$ because of $x\alpha - \Psi(x)\beta = y\beta \neq 0$, $y \in L$, then $x\alpha = (\Psi(x) + y)\beta$ and consequently $(\psi(x) + y)\beta = \psi(x)\beta$. Therefore we have $S^1(L) \supsetneqq S^1(M_2\beta \oplus V) = S^1(M_2\beta) \oplus S^1(V) \neq S^1(M_2\beta)$. This completes the proof[4]).

§ 2. Reduction of the problem

This section will be devoted to the reduction of two necessary conditions for SRCRT rings.

Assume that A is a SRCRT ring (for the definition cf. Introduction) and that $N(A)$ is the radical of A. If Q is an injective right A-module whose S^1-part is a finitely generated faithful $A/N(A)$-module, then Q is decomposed into a direct sum of finite number of indecomposable right A-modules. Let Q_0 be an indecomposable summand of Q. Now we shall show that Q_0 is quasi-primitive. Let P be a quasi-primitive right A-submodule of Q_0 such that the composition length of P is largest among one of all quasi-primitive right A-submodules of Q_0. Assume that Q_0 is not quasi-primitive, then there exists an element u of Q such that $P + uA \neq P$. Since Q_0 is injective and indecomposable, $S^1(Q_0)$ is simple and consequently $S^1(P + uA)$ is simple. Hence $P + uA$ is indecomposable and finitely generated. Since A is a ring of SRCRT, $P + uA$ is quasi-primitive. This contradicts our assumption. Thus Q_0 is quasi-primitive. Hence Q is a finitely generated right A-module. As was shown in [7, Theorems 2.5, 4.1 and 4.2] the duality holds between the caregories $\mathfrak{A}$ and $\mathfrak{B}$ consisting of all finitely generated right A-modules and left B-modules respectively, where B is the inverse A-homomorphism ring of Q, and it is known that B satisfies the minimum condition on left ideals. By Theorem 2.5 of [7] we have

PROPOSITION 2.1. *If a ring A is of SRCRT, then every finitely generated indecomposable left B-module has a unique minimal submodule.*

Thus it is not too much to say that the main object in this paper is to determine the complete structure of rings with the property stated in Proposition 2.1 in place of rings of SRCRT.

[4]) The lemma was first proved by Y. KAWADA, but this proof is an elaboration of his proof. In fact it will be known by this proof that the lemma holds for a case where the desending chain condition is assumed on M_1 and M_2 but not assumed on A.

PROPOSITION 2.2. *Under the same assumption of Proposition 2.1 B is a left generalized uni-serial ring.*

Proof. Let $Be_\varkappa$ be a primitive left ideal of B and $N(B)$ the radical of B. Since $Be_\varkappa/N(B)^\varrho e_\varkappa$ is indecomposable for every integer ϱ, it follows from Proposition 2.1 that $N(B)^{\varrho-1}e_\varkappa/N(B)^\varrho e_\varkappa$ is simple. Thus $Be_\varkappa$ has a unique composition series.

In case where A is an algebra, as was shown in [7, § 3] we can take the dual representation module A^* of A as the Q above and then $B=A$. Hence we obtain

COROLLARY 2.3. *If A is an algebra of SRCRT, then A is left generalized uni-serial. In particular, if A is of SRCRT and SLCRT, then A is generalized uni-serial.*

PROPOSITION 2.4. *Let B be a left generalized uni-serial ring. If M_1 and M_2 are quasi-primitive left B-modules such that $M_1 \approx Be_\mu/N^{\varrho_1}e_\mu$ and $M_2 \approx Be_\varkappa/N^{\varrho_2}e_\varkappa$, and $\varrho_1 \geqq \varrho_2$, where N is the radical of B and e_μ, $e_\varkappa$ are primitive idempotents of B. If θ is an isomorphism: $S^1(M_2) \to S^1(M_1)$. Then the interlacing module of M_1 and M_2 obtained by using θ as lacing isomorphism is directly decomposable if and only if θ is extendable to a monomorphism of M_2 into M_1.*

Proof. Proof of "if" part. Let α and β be injections of M_1 and M_2 into $\mathrm{Int}_\theta(M_1, M_2)$ respectively. Assume that Θ is a monomorphism which is an extension of θ. Then we readily see that $\mathrm{Int}_\theta(M_1, M_2) = M_1 \oplus M'$, where $M' = \{x_2\beta - \Theta(x_2)\alpha \mid x_2 \in M_2\}$.

Proof of "only if" part. Since B is left generalized uniserial, we can assume that $S^1(M_1) \approx N^{\varrho_1-1}e_\mu/N^{\varrho_1}e_\mu \approx N^{\varrho_1-1}e_\varkappa/N^{\varrho_2}e_\varkappa \approx S^1(M_2) \approx Be_\nu/Ne_\nu$, where N denotes the radical of B and e_μ, $e_\varkappa$, e_ν are primitive idempotents ($\varkappa$, μ and ν are not necessarily different from others). Put $\mathrm{Int}_\theta(M_1, M_2) = L$. We shall first consider the case (i) where $\varrho_1 > \varrho_2$. If L is decomposable, then from the direct decomposition of the top constituent[5]) of L it follows that L is decomposed into two quasi-primitive left B-modules L_1 and L_2: $L = L_1 \oplus L_2$, when $L_1 \approx M_1 \approx Be_\mu/N^{\varrho_1}e_\mu$ and $L_2 \approx M_2/S^1(M_2) \approx Be_\varkappa/N^{\varrho_2-1}e_\varkappa$, because otherwise it holds that $L_1 \approx Be_\mu/N^{\varrho_1-1}e_\mu$ and $L_2 \approx Be_\varkappa/N^{\varrho_2}e_\varkappa$, and that consequently $N^{\varrho_1-1}L = 0$, but this is a contradiction. Let $t_1 (= e_\mu t_1)$ and $t_2 (= e_\varkappa t_2)$ be generators of M_1 and M_2 respectively, and α and β injections of M_1 and M_2 into L respectively. Put

$$t_2\beta = e_\varkappa x_1 + e_\varkappa x_2, \quad x_1 \in L_1 \text{ and } x_2 \in L_2, \tag{6}$$

then $N^{\varrho_2-1}e_\varkappa x_1 = N^{\varrho_2-1}e_\varkappa t_2\beta \neq 0$ and $N^{\varrho_2}e_\varkappa x_1 = 0$. Hence we have that $e_\varkappa x_1$ is contained in $N^{\varrho_1-\varrho_2}L_1$ but not in $N^{\varrho_1-\varrho_2+1}L_1$. This shows that $N^{\varrho_1-\varrho_2}t_1 \approx N^{\varrho_1-\varrho_2}L_1 \approx Be_\varkappa/N^{\varrho_2}e_\varkappa$. Now let $e_\varkappa t_1'$ be a generator of $N^{\varrho_1-\varrho_2}t_1$. Put

$$t_1'\alpha = e_\varkappa y_1 + e_\varkappa y_2, \quad y_1 \in L_1 \text{ and } y_2 \in L_2. \tag{7}$$

Since L_1 is quasi-primitive, $e_\varkappa x_1$ and $e_\varkappa y_1$ are both generators of $N^{\varrho_1-\varrho_2}L_1$.

[5]) By the top constituent of L we shall mean the quotient module L/NL.

Hence there exists an element $a=e_\varkappa a e_\varkappa \in e_\varkappa B e_\varkappa$ such that

$$a y_1 = x_1. \tag{8}$$

Multiplying a to (7) from the left side, we have

$$a e_\varkappa t_1' \alpha = a y_1 + a y_2, \quad \text{hence by (8)} \quad a e_\varkappa t_1' \alpha - x_1 = a y_2 \in L_2.$$

Hence from (6) we conclude that $e_\varkappa a e_\varkappa t_1' \alpha - e_\varkappa t_2 \beta \in L_2$. Hence

$$n e_\varkappa a e_\varkappa t_1' \alpha - n e_\varkappa t_2 \beta = 0 \quad \text{for all } n \text{ of } N^{\varrho_2 - 1}.$$

Therefore $\theta(n e_\varkappa t_2) = n e_\varkappa a e_\varkappa t_1'$. Now, since B is a left generalized uni-serial ring, we can define an isomorphism Θ of Bt_2 onto Bt_1' by putting

$$\Theta(x e_\varkappa t_2) = x e_\varkappa a e_\varkappa t_1' \quad \text{for all } x \in B.$$

Then clearly Θ is an extension of θ which is a monomorphism. This completes the proof of this lemma for the case (i).

Next we shall consider the case (ii) where $\varrho_1 = \varrho_2 = \varrho$ and $e_\varkappa \neq e_\lambda$, and show L is indecomposable. Assume that $L = L_1 \oplus L_2$ and $L_1 \approx B e_\mu / N^\varrho e_\mu$, $L_2 \approx B e_\varkappa / N^{\varrho-1} e_\varkappa$. Using the notations of case (i), we have

$$t_2 \beta = e_\varkappa x_1 + e_\varkappa x_2, \quad e_\varkappa x_1 \in L_1 \text{ and } e_\varkappa x_2 \in L_2.$$

Since $e_\varkappa \neq e_\lambda$, it follows that $e_\varkappa x_1 \in N L_1$ and $N^{\varrho_2 - 1} t_2 \beta = 0$. But this is a contradiction.

Therefore the case leaved to be proved is following.

Case (iii) $\varrho_1 = \varrho_2 (= \varrho)$ and $e_\varkappa = e_\mu$. It will be seen that the proof of the case (i) is too applicable to this case and that the lacing isomorphism θ is extendable to a monomorphism. Thus the proof is completed.

PROPOSITION 2.5. *Let B be a left generalized uni-serial ring. If M_1 and M_2 be quasi-primitive left B-modules, and if θ is an isomorphism; $S^1(M_1) \to S^1(M_2)$. If θ is neither maximal nor extendable to a monomorphism, then the interlacing of M_1 and M_2 obtained by using θ as lacing isomorphism is indecomposable and has the non-simple S^1-part.*

Proof. The proof follows easily from Proposition 2.4 and Lemma 1.3.

PROPOSITION 2.6. *Let B be a left generalized uni-serial ring. If every indecomposable left B-module has a unique minimal submodule, then every isomorphism or its inverse between socles of two quasi-primitive left B-modules is either maximal or extendable to a monomorphism.*

Remark. Let B be an algebra over a field P which has basis $e_1, e_2, e_3, e_4, u_1, u_2, v_1, v_2, w_1$ and is defined by the following multiplication table, where blankets denote usually the zero-element. Then B is a left generalized uni-seril algebra and $S^2(Be_1) \approx Be_3 \approx S^2(Be_2)$ but $Be_1 \not\approx Be_2$, where e_1, e_2 and e_3 are primitive elements. Hence this example shows that there exists a left

generalized uniserial algebra which does not satisfy the conclusion of Proposition 2.6 and hence is not of SRCRT.

	e_1	e_2	e_3	e_4	u_1	v_1	w_1	u_2	v_2
e_1	e_1								
e_2		e_2							
e_3			e_3		u_1	v_1			
e_4				e_4			w_1	u_2	v_2
u_1	u_1								
v_1		v_1							
w_1			w_1		u_2	v_2			
u_2	u_2								
v_2		v_2							

COROLLARY 2.7. *Under the assumption of Proposition 2.6, it holds that every quasi-primitive left B-module is uniquely determined up to monomorphism by S^2-part.*

§ 3. Further necessary conditions for SRCRT rings

In this section we shall seek further necessary conditions in order that B may have the property such that every finitely generated indecomposable left B-module has a unique simple submodule.

PROPOSITION 3.1. *Let B be a left generalized uni-serial ring such that $N^2=0$, where N denotes the radical of B. Let L_i, $i=1, 2, 3$, be left B-modules. We assume further for L_i, $i=1, 2, 3$, that*

(i) $L_1=Be_\varkappa$, $L_2=Be_\lambda$, $L_3=Be_\mu$ *and* $e_\varkappa \neq e_\mu$, $e_\varkappa \neq e_\lambda$, $e_\lambda \neq e_\mu$,

(ii) S^1*-parts of* L_i, $i=1, 2, 3$, *are isomorphic to the others by isomorphisms* $\theta_{\lambda\varkappa}: S^1(L_1)=Ne_\varkappa \to S^1(L_2)=Ne_\lambda$; $\theta_{\mu\varkappa}: S^1(L_1) \to S^1(L_3)$,

(iii) *if φ is an isomorphism*: $S^1(L_1) \to L_2 \oplus L_3$ *defined by putting*

$$\varphi(xe_\varkappa)=\theta_{\lambda\varkappa}(xe_\varkappa)+\theta_{\mu\varkappa}(xe_\varkappa), \tag{9}$$

where $xe_\varkappa \in Ne_\varkappa$ *and* $\theta_{\lambda\varkappa}(xe_\varkappa) \in L_2$, $\theta_{\mu\varkappa}(xe_\varkappa) \in L_3$.

Then the interlacing L of L_1 and $L_2 \oplus L_3$ obtained by using φ as the lacing isomorphism is indecomposable.

Proof. Let α and β be injections of $L_2 \oplus L_3$ and L_1 into L respectively. Since φ is maximal, it follows from Lemma 1.3 that $S^1(L)=S^1((L_2 \oplus L_3)\alpha)$. Therefore if L is decomposed, L is decomposed into two modules M_1 and M_2, i.e. $L=M_1 \oplus M_2$. Thus the following cases are possible:

(a) $M_1/NM_1 \approx \bar{B}\bar{e}_\varkappa \oplus \bar{B}\bar{e}_\lambda$ and $M_2/NM_2 \approx \bar{B}\bar{e}_\mu$.

(b) $M_1/NM_1 \approx \bar{B}\bar{e}_\varkappa \oplus \bar{B}\bar{e}_\mu$ and $M_2/NM_2 \approx \bar{B}\bar{e}_\lambda$,

(c) $M_1/NM_1 \approx \bar{B}\bar{e}_\lambda \oplus \bar{B}\bar{e}_\mu$ and $M_2/NM_2 \approx \bar{B}\bar{e}_\varkappa$.

In case (a) let $g_1 (=e_\varkappa g_1)$ and $g_2 (=e_\lambda g_2)$ be a system of generators of M_1. Then

$$g_1=(e_\varkappa x_2 a+e_\varkappa x_3 b)\alpha+e_\varkappa x_1 c\beta, \quad g_2=(e_\lambda x_2' a+e_\lambda x_3' b)\alpha+e_\lambda x_1' c\beta$$

where a, b and c are generators of L_1, L_2 and L_3 respectively, and $x_2, x_2' \in Be_\lambda$, $x_3, x_3' \in Be_\mu$, and $x_1, x_1' \in Be_\varkappa$. Then $Ng_1 = Ne_\varkappa x_1 c\beta$ and $Ng_2 = Ne_\lambda x_2' a\alpha$. From (9) We have $Ng_1 \cap Ng_2 = 0$. This contradicts to $Ng_1 \cap Ng_2 = S(M_1)$.

We can similarly prove that the case (b) does not occur.

For the case (c) let $g_1 (= e_\lambda g_1)$ and $g_2 (= e_\mu g_2)$ be a system of generators of M_1. Then we have

$$g_1 = (e_\lambda x_2 a + e_\lambda x_3 b)\alpha + e_\lambda x_1 c\beta, \qquad g_2 = (e_\mu x_2' a + e_\mu x_3' b)\alpha + e_\mu x_1' c\beta,$$

where $x_2, x_2' \in Be_\lambda$, $x_3, x_3' \in Be_\mu$ and $x_1, x_1' \in Be_\varkappa$. Then we have $Ng_1 = Ne_\lambda x_2 a\alpha$, $Ng_2 = Ne_\mu x_3' b\alpha$. Hence $Ng_1 \cap Ng_2 = 0$. This is also a contradiction.

We say that quasi-primitive left B-modules M_1 and M_2 belong to the same S^2-class if $S^2(M_1) \approx S^2(M_2)$. The S^1-part of a quasi-primitive left B-module L is called also the S^1-part of S^2-class which contains L. Then by Corollary 2.7 we obtain

PROPOSITION 3.2. *Let B be a ring with the property that every indecomposable B-module has a unique minimal submodule. Then there exist at most two S^2-classes of quasi-primitive left B-modules whose S^1-part are isomorphic to the same simple left B-module.*

Let B be a left generalized uni-serial ring. Let P be a quasi-primitive left B-module such that $P \approx Be_\lambda / N^\varrho e_\lambda$ and $S^1(P) \approx N^{\varrho-1}e_\lambda / N^\varrho e_\lambda \approx Be_\varkappa / Ne_\varkappa$, $\varrho \geqq 2$. Let $t (= e_\lambda t)$ and $u (= e_\varkappa u)$ be generators of P and $N^{\varrho-1}e_\lambda$ respectively. Then for every element b of $e_\lambda B e_\lambda$ there corresponds an element b^* of $e_\varkappa B e_\varkappa$ such that

$$e_\varkappa b^* e_\varkappa u e_\lambda t = e_\varkappa u e_\lambda b e_\lambda t. \tag{10}$$

This correspondence is of course not well defined, but by taking the residue classes of element b and b^* modulo N we obtain an isomorphism of $\bar{e}_\lambda \bar{B} \bar{e}_\lambda$ into $\bar{e}_\varkappa \bar{B} \bar{e}_\varkappa$. Now we shall denote the image of $\bar{e}_\lambda \bar{B} \bar{e}_\lambda$ by $\mathfrak{R}^*$ and we put $\bar{e}_\varkappa \bar{B} \bar{e}_\varkappa = \mathfrak{Q}^*$.

On the other hand, for any elements b^* and b of $e_\varkappa B e_\varkappa$ and $e_\lambda B e_\lambda$ respectively, we can define automorphisms[6] φ_b and φ_{b^*} of P and $S^1(P)$ by putting

$$\varphi_b(x e_\lambda t) = x e_\lambda b e_\lambda t, \qquad \varphi_{b^*}(x' e_\varkappa u e_\lambda t) = x' e_\varkappa b^* e_\varkappa u e_\lambda t$$

for $x, x' \in B$. Then (10) implies

$$\varphi_b(e_\varkappa u e_\lambda t) = \varphi_{b^*}(e_\lambda u e_\lambda t). \tag{11}$$

Therefore if $\mathfrak{Q}^* = \mathfrak{R}^*$, every automorphism of $S^1(P)$ can be extendable to an automorphism of P.

Now we assume that B is self-basic[7] and $N^2 = 0$. Let Ne_λ be isomorphic to $\bar{B}\bar{e}_\varkappa (= Be_\varkappa / Ne_\varkappa)$ and $u (= e_\varkappa u e_\lambda)$ a generator of Ne_λ, i.e. $Ne_\lambda = e_\varkappa B e_\varkappa u e_\lambda$. Then in place of (10) we have

$$e_\varkappa b^* e_\varkappa u = u e_\lambda b e_\lambda \quad \text{for } b^* \in e_\varkappa B e_\varkappa,\ b \in e_\lambda B e_\lambda. \tag{10}'$$

[6] In this case $b \neq 0$, $b^* \neq 0$ (mod N).

[7] A ring A is said to be self-basic if its basic ring coincides with A itself.

$\mathfrak{Q}^*$ is considered as a right vector space over the quasi-field $\mathfrak{R}^*$ and let us denote by $(\mathfrak{Q}^*:\mathfrak{R}_r^*)$ the dimension of $\mathfrak{Q}^*$. We assume that $(\mathfrak{Q}^*:\mathfrak{R}_r^*)\geqq 2$, then $\mathfrak{Q}^*=\mathfrak{R}^*+\xi\mathfrak{R}^*+\eta\mathfrak{R}^*+\cdots$, where $\xi, \eta, \ldots$ are basis of $\mathfrak{Q}^*$ over $\mathfrak{R}^*$ and further assume that there exists another primitive ideal Be_μ, $\mu\neq\lambda$, whose S^1-part is isomorphic to $\overline{B}\,\overline{e}_\varkappa$. Let v be a generator of Ne_μ, i.e. $Ne_\mu=e_\varkappa Be_\varkappa v$, and put that $2\cdot Be_\lambda=Be_\lambda a\oplus Be_\lambda b$, where $Be_\lambda a\approx Be_\lambda b$ by isomorphisms: $be_\lambda a\leftrightarrow be_\lambda\leftrightarrow be_\lambda b$ for all b of B. Then we can define an isomorphism φ of Ne_μ into $2Ne_\lambda$ by putting

$$\varphi(e_\varkappa c e_\varkappa v)=e_\varkappa c e_\varkappa u a+e_\varkappa c e_\varkappa\xi u b, \tag{12}$$

where $e_\varkappa c e_\varkappa\in B$ and $2Ne_\lambda=Ne_\lambda a\oplus Ne_\lambda b$. Then the interlacing L of Be_μ and $2Be_\lambda$ obtained by using φ as lacing isomorphism is indecomposable.

To prove this assume that L is decomposable. Since φ is maximal, it follows by Lemma 1.3 that S^1-part of L is a direct sum of two simple modules. Thus L must be decomposed into two direct summands: $L=M_1\oplus M_2$. Then the following decompositions are possible:

(i) $M_1/NM_1\approx\overline{B}\,\overline{e}_\lambda\oplus\overline{B}\,\overline{e}_\mu$, $\quad M_2\approx\overline{B}\,\overline{e}_\lambda$,

(ii) $M_1/NM_1\approx 2\overline{B}\,\overline{e}_\lambda$, $\quad M_2\approx\overline{B}\,\overline{e}_\mu$.

We shall consider the case (i). Let $e_\lambda g_1$ and $e_\mu g_2$ be the system of generators of M_1, then

$$e_\lambda g_1=(e_\lambda x_1 e_\lambda a+e_\lambda x_2 e_\lambda b)\alpha+e_\lambda x_3 e_\mu\beta,$$
$$e_\mu g_2=(e_\mu z_1 e_\lambda a+e_\mu z_2 e_\lambda b)\alpha+e_\mu z_3 e_\mu\beta,$$

where $x_i, z_i\in B$ and α,β are injections $2Be_\lambda\to L$ and $Be_\mu\to L$ respectively. By (10)′ it follows that $e_\varkappa u e_\lambda g_1=e_\varkappa u(e_\lambda x_1 e_\lambda a+e_\lambda x_2 e_\lambda b)\alpha=(e_\varkappa x_1^* e_\varkappa u e_\lambda a+e_\varkappa x_2^* e_\varkappa u e_\lambda b)\alpha$ and by (12) we have $e_\varkappa v e_\mu g_2=e_\varkappa v(e_\mu z_3 e_\mu)\beta=e_\varkappa z_3^* e_\varkappa v e_\mu\beta=e_\varkappa z_3^* e_\varkappa(u e_\lambda a+\xi u e_\lambda b)\alpha$. Since $S^1(M_1)=Ng_1=Ng_2$, there exists an element $e_\lambda x e_\lambda\in B$ such that

$$\overline{e}_\varkappa\overline{x}_1^*\overline{e}_\varkappa=\overline{e}_\varkappa\overline{x}\,\overline{e}_\varkappa\overline{e}_\varkappa\overline{z}_3^*\overline{e}_\varkappa, \tag{13}$$

$$\overline{e}_\varkappa\overline{x}_2^*\overline{e}_\varkappa=\overline{e}_\varkappa\overline{x}\,\overline{e}_\varkappa\overline{e}_\varkappa\overline{z}_3^*\overline{e}_\varkappa\xi. \tag{14}$$

As $\overline{e}_\varkappa\overline{x}_1^*\overline{e}_\varkappa$ and $\overline{e}_\varkappa\overline{z}_3^*\overline{e}_\varkappa$ belong to $\mathfrak{R}^*$, it follows by (13) that $\overline{e}_\varkappa\overline{x}\,\overline{e}_\varkappa\in\mathfrak{R}^*$ and then by (14) we have that $\xi\in\mathfrak{R}^*$. But this is a contradiction. Hence the case does not occur.

Now we shall consider the case (ii). To this case we similarly denote by $e_\lambda g_1$ and $e_\lambda g_2$ the system of generators of M_1 and let $e_\lambda g_1=(e_\lambda x_1 e_\lambda a+e_\lambda x_2 e_\lambda b)\alpha+e_\lambda x_3 e_\mu\beta$, $e_\lambda g_2=(e_\lambda y_1 e_\lambda a+e_\lambda y_2 e_\lambda b)\alpha+e_\lambda y_3 e_\mu\beta$. Then $e_\varkappa u e_\lambda g_1=(\overline{e}_\varkappa\overline{x}_1^*\overline{e}_\varkappa u e_\lambda a+\overline{e}_\varkappa\overline{x}_2^*\overline{e}_\varkappa u e_\lambda b)\alpha$ and $e_\varkappa u e_\lambda g_2=(\overline{e}_\varkappa\overline{y}_1^*\overline{e}_\varkappa u e_\lambda a+\overline{e}_\varkappa\overline{y}_2^*\overline{e}_\varkappa u e_\lambda b)\alpha$, where $x_i^*, y_i^*, i=1,2$ are elements of B. By $Ng_1\cap Ng_2=S^1(M_1)$ there exists an element $\overline{e}_\varkappa\overline{x}\,\overline{e}_\varkappa$ of Q^* such that $\overline{e}_\varkappa\overline{x}_1^*\overline{e}_\varkappa=\overline{e}_\varkappa\overline{x}\,\overline{e}_\varkappa\cdot\overline{e}_\varkappa\overline{y}_1^*\overline{e}_\varkappa$, and $\overline{e}_\varkappa\overline{x}_2^*\overline{e}_\varkappa=\overline{e}_\varkappa\overline{x}\,\overline{e}_\varkappa\overline{e}_\varkappa\overline{y}_2^*\overline{e}_\varkappa$. Thus, $\overline{e}_\varkappa\overline{x}\,\overline{e}_\varkappa\in R^*$ and hence there exists a regular element $e_\lambda x' e_\lambda$ of $e_\lambda Be_\lambda$ such that $e_\varkappa x e_\varkappa u=u e_\lambda x' e_\lambda$. Then by (10)′ we have

$$\overline{e}_\lambda\overline{y}_1\overline{e}_\lambda=(\overline{e}_\lambda\overline{x}'\overline{e}_\lambda)^{-1}\overline{e}_\lambda\overline{x}_1\overline{e}_\lambda, \quad\text{and}\quad \overline{e}_\lambda\overline{y}_2\overline{e}_\lambda=(\overline{e}_\lambda\overline{x}'\overline{e}_\lambda)^{-1}\overline{e}_\lambda\overline{x}_2\overline{e}_\lambda.$$

Hence $e_\varkappa g_2 \equiv (e_\varkappa x' e_\varkappa)^{-1} e_\varkappa g_1 \bmod. NM_1$. Thus the top constituent of M_1 is simple and this contradicts to the decomposition (ii). This completes the proof and we have

PROPOSITION 3.3. *Let B be a self-basic generalized uni-serial ring with the radical N such that $N^2=0$. If B has the property stated in Proposition 2.1 and if the number of composition modules of $e_\varkappa N$ which are isomorphic to $e_\lambda B/e_\lambda N$ is not less than two, then there exists only one S^2-class of quasi-primitive modules whose S^1-part is isomorphic to $Be_\varkappa/Ne_\varkappa$.*

If B_0 is a basic ring of B, then following two covariant functors T_1 and T_2 compose an isomorphism in the sense of K. MORITA [7], where $T_1(M) = B_0 \underset{B_0}{\otimes} M$ for a left B-module M and $T_2(M_0) = B \underset{B_0}{\otimes} M_0$ for a left B_0-module M_0. Hence we can say that B has the property stated in Proposition 2.1 if and only if the basic ring B_0 of B has the same property.

PROPOSITION 3.4. *If B has the property such that every indecomposable left B-module has a unique minimal submodule and if the number of composition modules of $e_\varkappa N/e_\varkappa N^2$ which are isomorphic to $e_\lambda B/e_\lambda N$ is not less than two, then there exists only one S^2-class of quasi-primitive modules whose S^1-part is isomorphic to $Be_\varkappa/Ne_\varkappa$.*

Next under the first assumption of Proposition 3.3 we shall consider the case where $(\mathfrak{Q}^*:\mathfrak{R}_r^*) \geqq 3$, i.e.

$$\mathfrak{Q}^* = \mathfrak{R}^* + \xi_1 \mathfrak{R}^* + \xi_2 \mathfrak{R}^* + \cdots . \tag{15}$$

Let φ be an isomorphism of $Ne_\lambda c$ into $Ne_\lambda a \oplus Ne_\lambda b$ $(\approx 2 \cdot Ne_\lambda)$ obtained by putting

$$\varphi(xuc) = x\xi_1 ua + x\xi_2 ub, \tag{16}$$

where $Be_\lambda c \approx Be_\lambda a \approx Be_\lambda b \approx Be_\lambda$ by the following correspondences: $xe_\lambda c \leftrightarrow xe_\lambda a \leftrightarrow xe_\lambda b \leftrightarrow xe_\lambda$ for all $x \in B$, and u is a generator of Ne_λ. Let L be an interlacing of $Be_\lambda c$ and $Be_\lambda a \oplus Be_\lambda b$ obtained by using φ as the lacing isomorphism. Then L is indecomposable.

To prove this we shall first show that φ is maximal. Assume that φ is extendable and let Φ be an extension of φ, then we have $\Phi(e_\lambda c) = (e_\lambda z_1 a + e_\lambda z_2 b)$ for z_1 and z_2 of Be_λ. Hence by (10)$'$ $\Phi(ue_\lambda c) = \varphi(ue_\lambda c) = u(e_\lambda z_1 e_\lambda a + e_\lambda z_2 e_\lambda b) = \bar{e}_\varkappa \bar{z}_1^* \bar{e}_\varkappa ua + \bar{e}_\varkappa \bar{z}_2^* \bar{e}_\varkappa ub$, where $\bar{e}_\varkappa \bar{z}_i^* \bar{e}_\varkappa \in \mathfrak{R}^*$, $i=1, 2$. Then it follows from (16) that $\bar{e}_\varkappa \bar{z}_1^* \bar{e}_\varkappa = \xi_1$ and $\bar{e}_\varkappa \bar{z}_2^* \bar{e}_\varkappa = \xi_2$, but this contradicts to (15). Thus φ is maximal. Hence by Lemma 1.3 the S^1-part of M is a direct sum of two simple modules. And if L is decomposed, L must be a direct sum of two indecomposable modules M_1 and M_2, where we may assume $M_1/NM_1 \approx 2 \cdot \bar{B}\bar{e}_\lambda$. Let $e_\lambda g_1$ and $e_\lambda g_2$ be a system of generators of M_1, then we have

$$\left\{ \begin{array}{l} e_\lambda g_1 = (e_\lambda x_1 a + e_\lambda x_2 b)\alpha + e_\lambda x_3 c \beta \\ \text{and} \\ e_\lambda g_2 = (e_\lambda y_1 a + e_\lambda y_2 b)\alpha + e_\lambda y_3 c \beta , \end{array} \right. \tag{17}$$

where α and β are injections of $Be_\lambda a + Be_\lambda b$ and $Be_\lambda c$ into L respectively and x_i, y_i, $i=1, 2, 3$, are elements of B. It holds from (10)′ and (16) that

$$e_\varkappa u g_1 = \{(\bar{e}_\varkappa \bar{x}_1^* \bar{e}_\varkappa + \bar{e}_\varkappa \bar{x}_3^* \bar{e}_\varkappa \xi_1)\, u a + (\bar{e}_\varkappa \bar{x}_2^* \bar{e}_\varkappa + \bar{e}_\varkappa \bar{x}_3^* \bar{e}_\varkappa \xi_2)\, u b\}\alpha$$

and

$$e_\varkappa u g_2 = \{(\bar{e}_\varkappa \bar{y}_1^* \bar{e}_\varkappa + \bar{e}_\varkappa \bar{y}_3^* \bar{e}_\varkappa \xi_1)\, u a + (\bar{e}_\varkappa \bar{y}_2^* \bar{e}_\varkappa + \bar{e}_\varkappa \bar{y}_3^* \bar{e}_\varkappa \xi_2)\, u b\}\alpha,$$

where $\bar{x}_i^*$ and $\bar{y}_i^*$ are elements of $\mathfrak{R}^*$. Then, since $S^1(M_1)$ is simple, there exists an element $e_\lambda x e_\lambda$ of $e_\lambda B e_\lambda$ such that

$$(18)\quad \begin{cases} (e_\varkappa x_1^* e_\varkappa + e_\varkappa x_3^* e_\varkappa \xi_1)\, u = e_\varkappa x e_\varkappa (e_\varkappa y_1^* e_\varkappa + e_\varkappa y_3^* e_\varkappa \xi_1)\, u \\ \text{and} \\ (e_\varkappa x_2^* e_\varkappa + e_\varkappa x_3^* e_\varkappa \xi_2)\, u = e_\varkappa x e_\varkappa (e_\varkappa y_2^* e_\varkappa + e_\varkappa y_3^* e_\varkappa \xi_2)\, u, \end{cases}$$

here we can assume that either $\bar{e}_\varkappa \bar{x}_3^* \bar{e}_\varkappa$ or $\bar{e}_\varkappa \bar{y}_3^* \bar{e}_\varkappa$ is non-zero element of $\bar{e}_\lambda \bar{B} \bar{e}_\lambda$, for otherwise (17) and (18) induce that $M_1 \approx Be_\lambda$. Let

$$(19)\quad \bar{e}_\lambda \bar{x}_3^* \bar{e}_\lambda \neq 0.$$

Now consider a left vector space V over $\mathfrak{R}^*$ generated by colum matrices

$$v_1 = \begin{pmatrix} \tilde{x}_1 + \tilde{x}_3 \xi_1 \\ \tilde{x}_2 + \tilde{x}_3 \xi_2 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} \tilde{y}_1 + \tilde{y}_3 \xi_1 \\ \tilde{y}_2 + \tilde{y}_3 \xi_2 \end{pmatrix}$$

where $\tilde{x}_i, \tilde{y}_i$, $i=1, 2, 3$ means respectively $\bar{x}_i^*, \bar{y}_i^*$. Then (17) implies that

$$(\mathfrak{Q}^* \underset{\mathfrak{R}^*}{\otimes} V : \mathfrak{Q}_l^*) = 1,$$

where $\underset{\mathfrak{R}^*}{\otimes}$ means the tensor product over $\mathfrak{R}^*$ and the left term of equation means the left dimension of $\mathfrak{Q}^* \underset{\mathfrak{R}^*}{\otimes} V$ over $\mathfrak{Q}^*$. We shall alter the system of generator of V. By the assumption (18) we first have

$$V = \left\{ \begin{pmatrix} \tilde{x}_1 + \tilde{x}_3 \xi_1 \\ \tilde{x}_2 + \tilde{x}_3 \xi_2 \end{pmatrix}, \quad \begin{pmatrix} \tilde{y}_1' \\ \tilde{y}_2' \end{pmatrix} \right\}, \quad \text{where} \quad \tilde{y}_i' = \tilde{y}_i - \tilde{y}_3 \tilde{x}_3^{-1} \tilde{x}_i \quad \text{for } i = 1, 2.$$

In this case by the following reason we can assume $\tilde{y}_1' \neq 0$. Assume that $\tilde{y}_1' = 0$, then $\tilde{y}_2' = 0$, since $(\mathfrak{Q}^* \underset{\mathfrak{R}^*}{\otimes} V : \mathfrak{Q}_l^*) = 1$. And it holds that

$$(\bar{e}_\lambda \tilde{x}_1 \bar{e}_\lambda + \bar{e}_\lambda \tilde{x}_3 \bar{e}_\lambda \xi_1) = \bar{e}_\lambda \tilde{x}' \bar{e}_\lambda (\bar{e}_\lambda \tilde{y}_1 \bar{e}_\lambda + \bar{e}_\lambda \tilde{y}_3 \bar{e}_\lambda \xi_1),$$
$$\bar{e}_\lambda \tilde{x}_2 \bar{e}_\lambda + \bar{e}_\lambda \tilde{x}_3 \bar{e}_\lambda \xi_2 = \bar{e}_\lambda \tilde{x}' \bar{e}_\lambda (\bar{e}_\lambda \tilde{y}_2 \bar{e}_\lambda + \bar{e}_\lambda \tilde{y}_3 \bar{e}_\lambda \xi_2),$$

where[8]) $\bar{e}_\lambda \tilde{x}' \bar{e}_\lambda = \tilde{x}_3 \tilde{y}_3^{-1}$. Hence from (15) and (17) we have that $Bg_1 = Bg_2$ and this is a contradiction. Hence we obtain

$$V = \left\{ \begin{pmatrix} \tilde{x}_3 \xi_1 \\ \tilde{x}_2' + \tilde{x}_3 \xi_2 \end{pmatrix}, \quad \begin{pmatrix} \tilde{y}_1' \\ \tilde{y}_2' \end{pmatrix} \right\}, \quad \text{where} \quad \tilde{x}_2' = \tilde{x}_2 - \tilde{x}_3 \xi_1 \tilde{y}_1'^{-1} \tilde{y}_2'.$$

Similarly we have

$$V = \left\{ \begin{pmatrix} 0 \\ \tilde{x}_2' + \tilde{x}_3 \xi_2 - \tilde{x}_3 \xi_1 \tilde{y}_1'^{-1} \tilde{y}_2' \end{pmatrix}, \quad \begin{pmatrix} \tilde{y}_1' \\ \tilde{y}_2' \end{pmatrix} \right\}.$$

[8]) Assume $\tilde{y}_3 = 0$, then $\tilde{y}_1 = \tilde{y}_2 = 0$. Hence $\tilde{y}_3 \neq 0$.

Since $\tilde{y}_1 \neq 0$ and $(\mathfrak{L}^* \underset{\mathfrak{R}^*}{\otimes} V : \mathfrak{L}_i^*) = 1$, it must hold that $\tilde{x}_2' + \tilde{x}_3\xi_2 - \tilde{x}_3\xi_1\tilde{y}_1'^{-1}\tilde{y}_2' = 0$. By (19) we have $\tilde{x}_3^{-1}\tilde{x}_2 + \xi_2 + \xi_1\tilde{y}_1'\tilde{y}_2' = 0$, where $\tilde{x}_3^{-1}\tilde{x}_2, \tilde{y}_1'^{-1}\tilde{y}_2' \in \mathfrak{R}^*$. But ξ_i, $i = 1, 2$ are basis of $\mathfrak{L}^*$ over $\mathfrak{R}^*$ and this contradicts to $(\mathfrak{L}^* : \mathfrak{R}_l^*) \geqq 3$. Hence L is not decomposable. Thus we have

PROPOSITION 3.5. *If ring B satisfies the property stated in Proposition 2.1, the number of composition spaces of $e_\varkappa N/e_\varkappa N^2$ which are isomorphic to $e_\lambda B/e_\lambda N$ is at most 2, i.e.* $(e_\varkappa N e_\lambda/e_\varkappa N^2 e_\lambda : e_\lambda B e_\lambda/e_\lambda N e_\lambda) \leqq 2$.

Now we shall sum up the results of §§ 2 and 3.

PROPOSITION 3.6. *In order that B may be a ring possessing a property such that every finitely generated indecomposable left B-module has a unique minimal submodule, it is necessary that the following all conditions are satisfied:*

I. *B is a left generalized uni-serial ring, i.e. every primitive left ideal has a unique composition series.*

II. *Every isomorphism or its inverse between socles of two quasi-primitive left B-modules is either maximal or extendable to a monomorphism.*

III. *There exist at most two S^2-classes of quasi-primitive left B-modules whose S^1-parts are isomorphic to a given simple left B-module.*

IV. *The number of composition modules of $e_\varkappa N/e_\varkappa N^2$ which are isomorphic to $e_\lambda B/e_\lambda N$ is not larger than two, and if the number is equal to two, there exists only a S^2-class of quasi-primitive left B-modules such that its S^1-part is isomorphic to $Be_\varkappa/Ne_\varkappa$, where $e_\varkappa$ and e_λ are primitive idempotents of B.*

REMARK. It is known by examples that each condition of Proposition 3.6 is independent of all other conditions.

Assume that a ring B satisfies all conditions of Proposition 3.6. Let P be a quasi-primitive left B-module such that $P \approx Be_\lambda/N^\varrho e_\lambda$, $\varrho \geqq 2$ and $S^1(P) \approx N^{\varrho-1}e_\lambda/N^\varrho e_\lambda \approx Be_\varkappa/Ne_\varkappa$. Let $t\,(=e_\lambda t)$ and $u\,(=e_\varkappa u)$ be generators of P and $N^{\varrho-1}e_\lambda$ respectively and φ_i, $i = 1, 2, \ldots, \varrho$ automorphisms of $S^i(P)$ respectively. Then for every φ_i there corresponds uniquely an element $\bar{e}_\varkappa \bar{b}_i^* \bar{e}_\varkappa$ of $\bar{e}_\varkappa \overline{B}\, \bar{e}_\varkappa$ such that $\varphi_i(ut) = \bar{e}_\varkappa \bar{b}_i^* \bar{e}_\varkappa ut$. When we take all automorphism φ_i of $S^i(P)$ we shall denote the set of the corresponding elements $\bar{e}_\varkappa \bar{b}_i^* \bar{e}_\varkappa$ by $\mathfrak{E}_i^*$. Then ajoining the zero element of $\bar{e}_\varkappa \overline{B}\, \bar{e}_\varkappa$ to $\mathfrak{E}_i^*$, we have a quasi-field $\mathfrak{E}_i$. Clearly we have

$$\mathfrak{R}^* = {}_\varrho\mathfrak{E} \subseteqq \mathfrak{E}_{\varrho-1} \subseteqq \cdots \subseteqq \mathfrak{E}_2 \subseteqq \mathfrak{E}_1 = \mathfrak{L}^* \text{ [9]}$$

and according to the condition (ii) of Proposition 3.6 it holds that $\mathfrak{R}^* = \mathfrak{E}_\varrho = \mathfrak{E}_{\varrho-1} = \cdots = \mathfrak{E}_2$. Now we shall say that $\overline{B}\,\bar{e}_\varkappa$ is a simple module of the first kind (or belongs to the first kind) if $\mathfrak{R}^* = \mathfrak{L}^*$, and that $\overline{B}\,\bar{e}_\varkappa$ is the second kind if $\mathfrak{R}^* \neq \mathfrak{L}^*$. In the above definitions the kind of $\overline{B}\,\bar{e}_\varkappa$ does not depend on P for if $\mathfrak{R}^* \neq \mathfrak{L}^*$, i.e. $\mathfrak{E}_2 \neq \mathfrak{L}^*$, according to Proposition 3.3 there exists only one S^2-class of quasi-primitive modules whose S^1-part is isomorphic to $\overline{B}\,\bar{e}_\varkappa$.

§ 4. Preparations for the proof of sufficiency

Let B be a ring with a unit element 1. Let L be a left B-module and I a two-sided ideal of B. Then L is said to be an injective module relative

[9] For $\mathfrak{R}^*$ and $\mathfrak{L}^*$ see Proposition 3.3.

to I if the annihilator of I in L is injective as a left B/I-module. Dually a left B-module M is said to be a projective module relative to I if the quotient module M/IM is projective as a left B/I-module. Then we have immediately

LEMMA 4.1. *Let L be an injective (a projective) left B-module. Then L is injective (projective) relative to every two-sided ideal of B.*

Now we shall use the following notations.

Q: a finitely generated, injective and fully faithful left B-module;

A: the inverse B-homomorphism ring of Q, i.e. the ring of all B-endomorphisms of Q which is considered as a right operator domain of Q;

$(\ ,\ )_1$: a orthogonal pairing of B and Q defined by putting $(b, u)_1 = bu$ for $u \in Q$ and $b \in B$;

$(\ ,\ )_2$: a pairing of Q and A defined by putting $(u,a)_2 = ua$, for $u \in Q$ and $a \in A$;

I: a two-sided ideal of B;

Q_0: a submodule of Q consisting of all elements u such that u is annihilated by I in the pairing $(\ ,\ )_1$, i.e. $Q_0 = \mathbf{Ann}(I, Q)_1$.

J: an ideal of A consisting of all elements a such that a is annihilated by Q_0 in the pairing $(\ ,\ )_2$, i.e. $J = \mathrm{Ann}(Q_0; A)_2$. Then we shall prove

LEMMA 4.2. *Let L be a left B-module and R the Q-character module of L. If L_0 is a submodule of L consisting all elements which are annihilated by I, then the Q-character module of L_0 is isomorphic to a quotient right A-module R/RJ, more precisely $\mathrm{Ann}(L_0; R) = RJ$ with respect to the orthogonal pairing of L and R.*

Proof. According to the annihilator relations of character modules it is necessary and sufficient to prove the following equation

$$\mathrm{Ann}(I; L) = \mathrm{Ann}(RJ; L).$$

But $x \in \mathrm{Ann}(I; L) \Leftrightarrow ix = 0$ for all $i \in I \Leftrightarrow \varphi(ix) = 0$ for all $i \in I$ and $\varphi \in R = \mathrm{Char}_Q L \Leftrightarrow i \cdot \varphi(x) = 0$ for all $i \in I$ and $\varphi \in R \Leftrightarrow \varphi(x) \in Q_0$ for all $\varphi \in R \Leftrightarrow \varphi(x) \cdot j = 0$ for all $j \in J$ and $\varphi \in R$, because $Q_0 = \mathrm{Ann}(J; Q) \Leftrightarrow x \in \mathrm{Ann}(RJ; L)$. Thus the proof is completed.

Hence we have

THEOREM 4.3. *Using the same notations as in Lemma 4.2 if L is injective (projective) relative to I, then the Q-character module of L is projective (injective) relative to J of A.*

The following lemma will be used repeatedly in the later proofs.

LEMMA 4.4. *Let M and L be submodules contained in a left B-module. Assume that M is a direct sum of modules L_i, $i = 1, 2, \ldots, n$ and that*

$$x = \sum_{i=1}^{n} x_i \tag{20}$$

is a representation of an element x of $M \cap L$ as a sum of elements of L_i. Then we can define homomorphisms $\theta_i : M \cap L \to L_i$ by putting $\theta_i(x) = x_i$ for every

$x \in L \cap M$. Let I_0 be the set of all positive integers $i \leqq n$ such that θ_i are extendable to a homomorphism $\Theta_i : L \to L_i$. Then $M+L$ is decomposed into submodules $\sum_{i \in I_0} L_i$ and $\sum_{j \notin I_0} L_j + L'$, where $L' = \{a - \sum_{i \in I_0} \Theta_i(a) \mid a \in L\}$ and $\sum$ indicates "direct sum". If we define homomorphisms $\theta'_j : \left(\sum_{j \notin I_0} L_j\right) \cap L' \to L_j$ by putting $\theta'_j(x') = x'_j$ for all $x' \in \left(\sum_{j \notin I_0} L_j\right) \cap L'$, when $x' = \sum_{j \notin I_0} x'_j$, $x'_j \in L_j$, then all θ'_j are no more extendable to any homomorphisms $L' \to L_j$.

Proof. Denote by L' a left B-module consisting of all elements $a - \sum_{i \in I_0} \Theta_i(a)$, $a \in L$. Then we have clearly $L+M = L'+M = (L' + \sum_{j \notin I_0} L_j) + \sum_{i \in I_0} L_i$. If y is an elements of $(L' + \sum_{j \notin I_0} L_j) \cap \sum_{i \in I_0} L_i$, then $y = \sum_{i \in I_0} a_i = \sum_{j \notin I_0} a_j + (a - \sum_{k \in I_0} \Theta_k(a))$ for $a_i \in L_i$, $a \in L$, and hence $a = \sum_{i \in I_0} \{a_i + \Theta_i(a)\} - \sum_{j \notin I_0} a_j$. Then a is considered as an element of $M \cap L$. Therefore $\Theta_i(a) = \theta_i(a)$. Since M is a direct sum of modules L_i, $i = 1, 2, \ldots, n$ and $a = \sum_{i=1}^{n} \theta_i(a)$, we obtain $a_i + \theta_i(a) = \theta_i(a)$ for $i \in I$. Hence $a_i = 0$ for all $i \in I_0$ and consequently $y = 0$. The proof of later part of this lemma will be obtained easily and we shall omit the proof.

Hereafter we shall assume for all propositions in this section that the ring B satisfies the all conditions stated in Proposition 3.6.

PROPOSITION 4.5. *Let P_1 and P_2 be quasi-primitive left B-modules such that their S^1-parts are isomorphic to $\overline{B}\,\overline{e}_\mu$ and are of first kind. Suppose that P_1 and P_2 do not belong to the same S^2-class. Then every interlacing L of P_1 and P_2 obtained by using an isomorphism: $S^1(P_1) \to S^1(P_2)$ as lace isomorphism is injective relative to N^ϱ, where $\varrho = \mathrm{Min}[\varrho_1, \varrho_2]$ and $\varrho_i = [P_i]$ the composition length of P_i, $i = 1, 2$.*

Proof. Assume that $\varrho_1 \leqq \varrho_2$. According to Lemma 1.2 it suffices to prove that every interlacing of L and M is decomposable, where M is any quasi-primitive left B-module such that its S^1-part is isomorphic to $\overline{B}\,\overline{e}_\mu$ and $N^{\varrho_1} M = 0$. Let $S^2(P_1) \approx Be_\varkappa / N^2 e_\varkappa$, $S^2(P_2) \approx Be_\lambda / N^2 e_\lambda$, $S^2(M) \approx Be_\nu / Ne_\nu$ and $S^1(P_1) \approx S^1(P_2) \approx S^1(M) \approx Be_\mu / Ne_\mu$, then from the condition III of Proposition 3.6 we obtain that $S^2(M)$ is isomorphic to either $S^2(P_1)$ or $S^2(P_2)$. Now assume that $S^2(M) \approx S^2(P_1)$, then since $S^1(P_1)$ is of first kind every isomorphism $\theta : S^1(M) \to S^1(P_1)$ have an extension $\theta' : S^2(M) \to S^2(P_1)$. From the condition II of Proposition 3.6 we further have an extension $\Theta : M \to P_1$ of θ', and consequently an extension of θ, which is a monomorphism. Since the monomorphism Θ is considered as a monomorphism of M into L, by Lemma 4.4 $L+M$ is decomposed into L and M', where $M' = \{x - \Theta(x) \mid x \in M\}$.

COROLLARY 4.6. *Let a ring B satisfy all conditions of Proposition 3.6. Let P_i, $i = 1, 2, \ldots, n$ and P be quasi-primitive left B-modules contained in a left B-module such that $[P] \leqq [P_i]$, $i = 1, 2, \ldots, n$, where by $[P]$ and $[P_i]$ we denote the composition lengths of P and P_i respectively. If the sum M of P_i, $i = 1, 2, \ldots, n$ is a direct sum of P_i, and if $M \cap P \geqq S^2(P)$, then $M+P$ is*

decomposable and has at least an indecomposable summand which is projective relative to N^{ϱ}, where $\varrho=[P]$.

Proof. Since B is a left generalized uni-serial ring and P is quasi-primitive, $M\cap P$ is cyclic. Let u be a generator of $M\cap P$. Similarly as in Lemma 4.4 let

$$u=\sum_{i=1}^{n}u_i \tag{21}$$

be a representation of as a sum of elements of M and θ_i homomorphisms: $M\cap P\to P_i$ defined by putting $\theta_i(xu)=xu_i$ for $x\in B$. Since P_i are quasi-primitive and B is a generalized uni-serial ring, there exists at least a P_i such that $Bu_i\supseteqq S^2(P_i)$, then from the condition II of Proposition 3.6 θ_i is extendable to an homomorphism: $P\to P_i$. Hence the conclusion follows from Lemma 4.4.

REMARK. In Corollary 4.6 let us denote by I_0 the set consisting of all integer i such that $Bu_i\supseteqq S^2(P_i)$. Then $M+P=\sum_{i\in I_0}P_i+\left(\sum_{j\notin I_0}P+P'\right)$, where $P'=\left\{a-\sum_{i\in I_0}\Theta_i(a)\,\middle|\,a\in P \text{ and } \Theta_i \text{ are extensions of } \theta_i\right\}$. Let u' be a generator of $\left(\sum_{j\notin I_0}P_j\right)\cap P'$ and let

$$u'=\sum_{j\notin I_0}u'_j \tag{22}$$

be a representation of u' as an element of $\sum_{j\notin I_0}P_i$, then we can assume that $S^1(P_j)\supseteqq Bu'_j$ and $S^1(P')\supseteqq Bu'$.

PROPOSITION 4.7. *Let P_i, $i=1, 2, \ldots, n$, and P be quasi-primitive left B-modules contained in a module such that $S^1(P_i)\approx S^1(P)$, $S^2(P_i)\approx S^2(P_{i'})$, $S^2(P_i)\not\approx S^2(P)$, $1\leqq i, i'\leqq n$ and that $S^1(P)$ is of the first kind. If the sum of P_i, $i=1, 2, \ldots, n$ is a direct sum, then $M+P$ is decomposable and every indecomposable summand of $M+P$ is projective or injective relative to a suitable power of the radical N.*

Proof. When $n=1$, by Proposition 4.5 the proof is obvious. Thus we proceed the proof by induction. Let us denote by $[P_i]$ and $[P]$ the composition lengths of P_i and P respectively. Assume that $[P_i]\leqq[P_{i+1}]$ for $i=1, 2, \ldots, n-1$. Let u be a generator of $S^1(P)$ and let

$$u=\sum_{i=1}^{n}u_i \tag{23}$$

be the representation of u as a sum of elements of P_i, where $u_i\in S^1(M_i)$, $i=1, 2, \ldots, n$. If $u_i=0$ for some i, then from the assumption of induction the proof is obvious. Now assume that all u_i in (1) are not a zero element. Then we have

$$P_2+P_3+\cdots+P_n+P=P_2\oplus P_3\oplus\cdots\oplus P_n\oplus P, \tag{24}$$

where right hand of (24) implies the direct sum of P_2, P_3, P_{n-1}, P_n and P. From (23) we have

$$u_1=\sum_{i=2}^{n}(-u_i)+u. \tag{25}$$

According as Lemma 4.4 we shall define homomorphisms (in fact isomorphisms) $\theta_i: S^1(P_1) \to S^1(P_i)$, $i=2, \ldots, n$ by putting

$$\theta_i(x u_1) = -x u_i \qquad \text{for all } x \in B. \tag{26}$$

Since $\bar{B}\bar{e}_i$ is of the first kind, every θ_i, $i=1, 2, \ldots, n$, can be extended to a monomorphism Θ_i, and applying Lemma 4.4 to $P_2 \oplus P_3 \oplus \cdots \oplus P_n \oplus P$ and P_1, we have L is decomposed into modules $P_2 \oplus P_3 \oplus \cdots \oplus P_n$ and $P + P_1'$, where $P_1' = \{a - \sum_{i=2}^{n} \Theta_i(a) \mid a \in P_1\}$. Hence since $S^2(P) \neq S^2(P')$ it follows by Proposition 4.5 that $P + P'$ is injective relative to a power of the radical. This completes the proof.

PROPOSITION 4.8. *Let P_i, $i=1, 2, \ldots, n$, P_{n+j}, $j=1, 2, \ldots, m$ and P be quasi-primitive left B-modules contained in a modules such that $S^1(P_i) \approx S^1(P_{n+j}) \approx S^1(P)$, $S^2(P_i) \approx S^2(P_1)$, $S^2(P_{n+j}) \approx S^2(P_{n+1})$, $S^2(P_1) \neq S^2(P_{n+1})$ and $S^1(P)$ is of the first kind. If the sum M of P_i and P_{n+j}, $i=1, 2, \ldots, n$, $j=1, 2, \ldots, m$ is a direct sum of P_i and P_{n+j} and $M \cap P = S^1(P)$, then $M+P$ is decomposable and every indecomposable summand of $M+P$ is either projective or injective to a suitable power of N.*

Proof. From the condition III of Proposition 3.6, the S^2-class of P is either the S^2-class of P_i or P_{n+j}, so we shall assume $S^2(P) \approx S^2(P_i)$. We shall put that $M_1 = P_1 \oplus \cdots \oplus P_n$ and $M_2 = P_{n+1} \oplus \cdots \oplus P_{n+m}$.

To begin with we shall consider the case where $n=1$ and $m=1$. Let u be a generator of $S^1(P)$ and put

$$u = u_1 + u_2, \qquad u_1 \in S^1(P_1), \qquad u_2 \in S^1(P_2) \tag{27}$$

a representation of u as a sum of elements of $S^1(P_i)$. If[10] $u_1 = 0$, then $M+P = M_1 \oplus (M_2 + P)$ and it is known by Proposition 4.5 that $M_2 + P$ is injective relative to a power of the radical. When $u_1 \neq 0$ and $u_2 \neq 0$ in (27), according as Lemma 4.4 we shall define homomorphism $\theta: S^1(P) \to S^1(P_1)$ by putting $\theta(xu) = x u_1$ for all $x \in B$. Then θ is extendable, since $S^1(P)$ is of the first kind and it follows by Lemma 4.4 and Proposition 4.5 that $M+P$ is decomposed into a quasi-primitive module and injective module relative to a power of N.

Now we proceed the proof by double induction with respect to pairs (n, m). Assume that this proposition holds for all pairs (r', s'), where $r' \leqq r-1$ and $s' \leqq s$ or $r' \leqq r$ and $s' \leqq s-1$. We shall prove the proposition for the pair $(n=r, m=s)$. Let

$$u = \sum_{i=1}^{r} u_i + \sum_{j=r+1}^{r+s} u_j \tag{28}$$

be the representation of generator of $S^1(P)$ as an element of M. If u_i or u equals to zero element for some i or j, the proof will be reduced to a case $n \leqq r-1$, $m \leqq s$ or $n \leqq r$, $m \leqq s-1$. Thus we can assume that

$$u_i \neq 0 \quad \text{and} \quad u_j \neq 0 \quad \text{for all } i \text{ and } j. \tag{29}$$

[10] If $u_2 = 0$, then it is easily seen that every direct summand of $M+P$ is quasi-primitive, i.e. projective relative to a power of the radical.

Further without losing of generality we can assume that $[P_i] \leqq [P_{i+1}]$, $i=1, 2, \ldots, r-1$. We shall define homomorphisms θ_i; $S^1(P) \to S^1(P_i)$, $i=1, 2, \ldots, r$, by putting $\theta_i(xu) = xu_i$ for all $x \in B$. Since $S^1(P)$ is of the first kind, θ_i has an extension Θ_i and applying Lemma 4.4 we have a decomposition of $M = P_1 \oplus P_2 \oplus \cdots \oplus P_r \oplus (P' + M_2)$, $P' = \{x - \sum_{i=1}^{r} \Theta_i(x) | x \in P\}$. Therefore the proof is reduced to the proof of Proposition 4.6.

Next we shall prove similar propositions concerning interlacings of quasi-primitive modules of the second kind.

Let P be a quasi-primitive left B-module such that $P \approx Be_\lambda/N^\varrho e_\lambda$ and $S^1(P) \approx N^{\varrho-1}e_\lambda/N^\varrho e_\lambda \approx Be_\varkappa/Ne_\varkappa$, where ϱ is an integer and e_λ, $e_\varkappa$ are primitive idempotents of B. If $\bar{B}\bar{e}_\varkappa$ is a simple module of the second kind, using the same notations as those of § 3 we have

$$\mathfrak{Q}^* = \mathfrak{R}^* + \xi \mathfrak{R}^*. \tag{30}$$

Let P_a and P_b be quasi-primitive left B-modules which are isomorphic to P by the following correspondences $xa \leftrightarrow x \leftrightarrow xb$ for all $x \in P$. Then an isomorphism $\theta: S^1(P)b \to S^1(P)b$, defined by putting

$$\theta(cue_\lambda tb) = c\xi ue_\lambda ta \quad \text{for all } c \in B, \tag{31}$$

is maximal. Hence the interlacing M of P_a and P_b obtained by using θ as lacing isomorphism is indecomposable, for by Lemma 1.3 it follows that $S^1(M) = S^1(P_a)$ and $S^1(P_a)$ is simple. Now we shall further prove

PROPOSITION 4.9. *M is injective relative to N^ϱ.*

Proof. Let P' be a quasi-primitive left B-module whose S^1-part of P' is isomorphic to $\bar{B}\bar{e}_\varkappa$. If $[P'] < [P]$, P' is monomorphic to P, for $\bar{B}\bar{e}_\varkappa$ is of the second kind. Hence by Lemma 1.2 it is sufficient to show that every interlacing of P with M has a direct summand which is isomorphic to M, when by the condition (II) of Proposition 3.6 and Lemma 4.4 every lace isomorphism may be assumed to be defined between $S^1(P)$ and $S^1(M)$. Let φ be any lace isomorphism: $S^1(P) \to S^1(M)$, then there exist two elements x^* and y^* of B such that

$$\varphi(ce_\varkappa ue_\lambda t) = c(x^* + y^*\xi)uta\alpha, \tag{32}$$

where α is the injection of P_a into M. For x^* and y^* there exist two elements x and y of $e_\lambda Be_\lambda$ such that they satisfy the following equations

$$\begin{cases} e_\varkappa x^* e_\varkappa ue_\lambda \equiv e_\varkappa ue_\lambda xe_\lambda & (\text{mod. } N^\varrho) \\ e_\varkappa y^* e_\varkappa ue_\lambda \equiv e_\varkappa ue_\lambda ye_\lambda & (\text{mod. } N^\varrho) \end{cases} \tag{33}$$

We shall consider a homomorphism $\Phi: P \to M$ which maps the generator $e_\lambda t$ of P to an element $xe_\lambda ta\alpha + ye_\lambda tb\beta$, where β is the injection of P_b into M. Then Φ is an extension of φ, because by (31) und (33)

$$\begin{aligned} \Phi(ue_\lambda t) &= uxta\alpha + uytb\beta = x^* uta\alpha + y^* utb\beta \\ &= x^* uta\alpha + y^* \xi uta\alpha = (x^* + y^*\xi)uta\alpha. \end{aligned}$$

Hence by Lemma 4.4 M splits from Int $\varphi(P, M)$ as a direct summand. This completes the proof.

REMARK. Together with the result of Proposition 4.5 we can state generally that if P_1 and P_2 are two quasi-primitive modules such that the isomorphism $\vartheta: S^1(P_1) \to S^1(P_2)$ is maximal, then the interlacing of P_1 and P_2 obtained by using ϑ as lacing isomorphism is injective relative to N^ϱ, where N^ϱ annihilates neither P_1 nor P_2.

PROPOSITION 4.10. *Let M be a sum of the following left B-modules M_0 and P; M_0 is a direct sum of quasi-primitive left B-modules P_i, $i=1, 2, \ldots, n$ having as the S^1-part the same $\overline{B}\,\overline{e}_\lambda$ which belong to the second kind; P is a quasi-primitive left B-module having $\overline{B}\,\overline{e}_\lambda$ as the S^1-part. If $M_0 \cap P = S^1(P)$, and if the composition length of P is not longer than that of P_i, then M is directly decomposable and every direct summand is either projective or injective relative to a suitable power of the radical.*

Proof. When $n=1$, this proposition is reduced to Proposition 4.9. Hence we shall proceed the proof by induction for n. Assume that this proposition holds for $n \leqq r-1$. Put that $P_i \approx Be_{\varkappa_i}/N^{\varrho_i}e_{\varkappa_i}$, $P \approx Be_\varkappa/N^\varrho e_\varkappa$ and $N^{\varrho_i-1}e_{\varkappa_i}/N^{\varrho_i}e_{\varkappa_i} \approx N^{\varrho-1}e_\varkappa/N^\varrho e_\varkappa \approx \overline{B}\,\overline{e}_\lambda$, $i=1, 2, \ldots, r$, where $\varrho_1 \leqq \varrho_2 \leqq \cdots \leqq \varrho_r$. Let us denote by P_i', P_i'', $i=1, 2, \ldots, r$, the annihilators of N^{ϱ_i} and N^ϱ in P_i respectively. Since $\overline{B}\,\overline{e}_\lambda$ belongs to the second kind, P_i' are isomorphic to P_1 and P_i'' are isomorphic to P. Denote by $e_\varkappa t$, $e_{\varkappa_i}t_i$, $e_\varkappa s_i$, $e_\lambda u e_\varkappa$ and $e_\lambda v e_{\varkappa_i}$, $i=1, 2, \ldots, r$ the generators of P, P_i', P_i'', $N^{\varrho-1}e_\varkappa$ and $N^{\varrho_i-1}e_{\varkappa_i}$ respectively[11]). Then we have

$$e_\lambda u e_\varkappa t = \sum_{i=1}^{r} (a_i^* + \xi b_i^*)\, e_\lambda v e_{\varkappa_i} t_i, \tag{34}$$

where $a_i^*, b_i^* \in \mathfrak{R}^*$ (cf. (30)). Here every element b_i^* may be assumed to be non-zero elements, for if $b_i^*=0$ for some i, then an isomorphism $\theta: P \to P_i''$, defined by putting $\theta(ce_\lambda u e_\varkappa t) = ca_i^* e_\lambda v e_{\varkappa_i} t_i$, $c \in B$ is extendable and hence applying Lemma 4.4 we arrive at the case where $n=r-1$. Then there exist elements a_i and b_i, $i=1, 2, \ldots, r$ of $\overline{e}_\varkappa \overline{B}\,\overline{e}_\varkappa$ and $\overline{e}_{\varkappa_i}\overline{B}\,\overline{e}_{\varkappa_i}$ such that

$$a_i^* e_\lambda u e_\varkappa = e_\lambda u e_\varkappa a_i, \qquad b_i^* e_\lambda v e_{\varkappa_i} = e_\lambda v e_{\varkappa_i} b_i. \tag{35}$$

Put

$$e_\varkappa g = \sum_{i=1}^{r} a_i e_\varkappa s_i - t, \qquad e_{\varkappa_1} h = \sum_{i=1}^{r} b_i e_{\varkappa_i} t_i, \tag{36}$$

and consider quasi-primitive left B-modules $Be_\varkappa g$ and $Be_{\varkappa_1}h$. Then the generators of $S^1(Be_\varkappa g)$ and $S^1(Be_{\varkappa_1}h)$ are clearly $e_\lambda u e_\varkappa g$ and $e_\lambda v e_{\varkappa_1} h$ respectively. By (34), (35) and (36)

$$e_\lambda u e_\varkappa g = \sum_{i=1}^{r} e_\lambda u e_\varkappa a_i e_\varkappa s_i - e_\lambda u e_\varkappa t = \sum_{i=1}^{r} a_i^* e_\lambda u e_\varkappa s_i - \sum_{i=1}^{r} (a_i^* + \xi b_i^*)\, e_\lambda v e_{\varkappa_i} t_i$$

$$= \sum_{i=1}^{r} a_i^* e_\lambda v e_{\varkappa_i} t_i - \sum_{i=1}^{r} (a_i^* + \xi b_i^*)\, e_\lambda v e_{\varkappa_i} t_i = -\xi \Big(\sum_{i=1}^{r} b_i^* e_\lambda v e_{\varkappa_i} t_i \Big)$$

[11]) In this case we assume that $e_\varkappa t$, $e_{\varkappa_i}t_i$, $e_\varkappa s_i$, $e_\lambda u e_\varkappa$ and $e_\lambda v e_{\varkappa_i}$ are chosen as follows: If $e_\varkappa w e_{\varkappa_i}$ is a generator of $N^{\varrho_i-\varrho}e_{\varkappa_i}$, then $e_\varkappa v e_{\varkappa_i} = e_\lambda u \cdot w e_{\varkappa_i}$, and $e_\lambda u e_\varkappa s_i = e_\lambda v e_{\varkappa_i} t_i$.

and

$$e_\lambda v e_{\varkappa_1} h = \sum_{i=1}^{r} e_\lambda v e_{\varkappa_1} b_i^* e_{\varkappa_1} t_i = \sum_{i=1}^{r} b_i^* e_\lambda v e_{\varkappa_1} t_i .$$

Hence $e_\lambda u e_\varkappa g = -\xi e_\lambda v e_{\varkappa_1} h$ and consequently $S^1(Be_\varkappa g) = S^1(Be_{\varkappa_1} h)$. This implies from the remark of Proposition 4.9 that the sum of $Be_\varkappa g$ and $Be_{\varkappa_1} h$ is relative injective. Since all $b_i \neq 0$, by the assumption that $S^1(M) = S^1(M_0)$, we have

$$M = (Be_\varkappa g + Be_{\varkappa_1} h) \oplus P_2 \oplus \cdots \oplus P_r .$$

This completes the proof.

§ 5. The proof of sufficiency

We are now ready to prove that the all conditions obtained in Proposition 3.6 are sufficient for ring B to satisfy a property that every finitely generated indecomposable left B-module has a unique minimal submodule.

PROPOSITION 5.1. *Let B be a left generalized uni-serial ring. Let M be a left B-module and $S^1(M) = \overline{M}_1 \oplus \overline{M}_2 \oplus \cdots \oplus \overline{M}_t$ the ideal decomposition of the S^1-part of M, i.e. any two simple left B-modules of $S^1(M)$ are isomorphic to each other if and only if they are contained in the same direct summand M_i. Then we have a direct decomposition $M = M_1 \oplus M_2 \oplus \cdots \oplus M_t$ of such that $S^1(M_i) = \overline{M}_i$.*

Proof. In this proof all arguments in the proof of Proposition 5.2 will be assumed. This may be considered as a continuation of the proof of Proposition 5.2. We shall determine generators of a suitable system of generators of M from a given system of generators by the following successive way. Let $\{e_1 g_1, \ldots, e_n g_n\}$ be a given system of generators of M, S the submodule of M generated by $e_1 g_1, \ldots, e_r g_r$, $r < n$, and $[Be_i g_i] \geqq [Be_{i+1} g_{i+1}]$, $i \leqq n-1$. Assume that r generators $e_1 g_1, \ldots, e_r g_r$ are already choosed such as every direct summand D_j of S is isomorphic to some $Be_i g_i$ or $Be_{k+1} g_{k+1} + Be_k g_k$, $i, k+1 \leqq r$, where $Be_{k+1} g_{k+1} + Be_k g_k$ is indecomposable injective relative to N^ϱ, $\varrho = [Be_k g_k]$. If $S \cap Be_{r+1} g_{r+1} = 0$, we may take $e_{r+1} g_{r+1}$ as the $r+1$-th member of the suitable system of generators. But in case $S \cap Be_{r+1} g_{r+1} \neq 0$, we consider the following equation and homomorphisms $\theta_j : Be_{r+1} g_{r+1} \cap S \to D_j$; $x = x_1 + \cdots + x_s$, $x \in Be_{r+1} g_{r+1}$, $x_j \in D_j$ and $\theta_j(x) = x_j$. Then by Lemma 4.4 we can take $Be_{r+1} g'_{r+1} = \{x - \sum_j \Theta_j(x) \mid x \in Be_{r+1} g_{r+1} \cap S$ and the sum is taken over all $\Theta_j : Be_{r+1} g_{r+1} \to D_j$, which are extensions of θ_j respectively$\}$ in place of $Be_{r+1} g_{r+1}$. From the condition II of Proposition 3.6[12]) it follows that all images of the corresponding new homomorphisms $\theta'_j : S \cap Be_{r+1} g'_{r+1} \to D_j$ must be contained in $S^1(D_j)$. Hence θ'_j are isomorphisms or zero-mappings. If there exists an isomorphism θ_j, we take $e_{r+1} g'_{r+1}$, as the $r+1$-th member of the suitable system of generators. If all θ'_j are zero-mappings, we line up generators $e_{r+1} g'_{r+1}$, $e_{r+2} g_{r+2}, \ldots, e_n g_n$ in accordance with the composition lengths $[Be_i g_i]$, $i = r+1, \ldots, n$. And we repeat the same process untill we can determine the $r+1$-th member of the suitable system. Then it is known

[12]) Cf. Remark of Corollary 4.6.

from the result of Proposition 5.2 that, when we determine the $r+1$-th member, the first $r+1$ generators satisfy the assumption which was in the beginning supposed for the first r generators. Thus the conclusion of this proposition is clear, as soon as we determine all generators of the suitable system.

PROPOSITION 5.2. *Let B be a ring satisfying all conditions of Proposition 3.6. A left B-module M is indecomposable if and only if the S^1-part of M is simple.*

Proof. Let M be an indecomposable left B-module. We assume that every simple submodule of $S^1(M)$ is isomorphic to $\overline{B}\bar{e}_\lambda$. Now we divide the proof into two parts.

I. $\overline{B}\bar{e}_\lambda$ is of the first kind. Let $\{e_1g_1, e_2g_2, \ldots, e_ng_n\}$ be a system of generators of M and further

$$e_1g_1,\ e_2g_2, \ldots, e_ng_n \tag{37}$$

a sequence of e_ig_i, $i=1, 2, \ldots, n$, such that $[Be_ig_i] \geqq [Be_{i+1}g_{i+1}]$. Such a sequence of generators will be called a normal sequence of the system of generators. Take Be_1g_1 and Be_2g_2, then the following cases are possible

Ia. $Be_1g_1 \cap Be_2g_2 = 0$, i.e. $Be_1g_1 + Be_2g_2 = Be_1g_1 \oplus Be_2g_2$.

Ib. $Be_1g_1 \cap Be_2g_2 \neq 0$ and the S^2-classes of Be_1g_1 and Be_2g_2 are different each other.

Ic. $Be_1g_1 \cap Be_2g_2 \neq 0$ and the S^2-classes of Be_1g_1 and Be_2g_2 are identical.

If the case Ib occurs, then it follows by Proposition 4.5 that $Be_1g_1 + Be_2g_2$ is injective relative to N^{ϱ_2}, where $\varrho_2 = [Be_2g_2]$. Since $[Be_ig_i] \leqq \varrho_2$ for $i \geqq 3$ it follows that M is decomposable for $n \geqq 3$. Next if we have the case Ic, we shall consider the identity mapping θ: $Be_1g_1 \cap Be_2g_2 \to Be_1g_1 \cap Be_2g_2$. As $\overline{B}\bar{e}_\lambda$ is of the first kind, θ is extendable to a monomorphism $Be_2g_2 \to Be_1g_1$, and hence by Lemma 4.4 we obtain a decomposition;

$$Be_1g_1 + Be_2g_2 = Be_1g_1 \oplus Be_2g_2',$$

where it is seen $[Be_2g_2] > [Be_2g_2']$. Then we can take e_2g_2' together with e_1g_1, e_3g_3, e_4g_4, $\ldots$, e_ng_n as a system of generators of M. Now arrange the elements of the new system in the normal sequence. If for the new system of generators we have not the case Ib, as is shown above we meet the case Ia. In place of discussing on the case Ia we consider the more general case such that $Be_1g_1 + Be_2g_2 + \cdots + Be_rg_r = Be_1g_1 \oplus Be_2g_2 \oplus \cdots \oplus Be_rg_r$, where Be_ig_i, $i=1, 2, \ldots, r$, belong to the same S^2-class of quasi-primitive left B-modules. At the same time the following cases are possible.

Id. The S^2-classes of Be_1g_1 and $Be_{r+1}g_{r+1}$ are different each other and $U \cap Be_{r+1}g_{r+1} \neq 0$, where $U = Be_1g_1 + Be_2g_2 + \cdots + Be_rg_r$.

Ie. The S^2-classes of Be_1g_1 and $Be_{r+1}g_{r+1}$ are identical and $U \cap Be_{r+1}g_{r+1} \neq 0$.

For the case Id applying Proposition 4.7 we obtain $U + Be_{r+1}g_{r+1} = Be_1g_1 \oplus Be_2g_2 \oplus \cdots \oplus Be_{r+1}g_{r+1} \oplus Q$, where Q is injective relative to $N^{\varrho_{r+1}}$ and $\varrho_{r+1} = [Be_{r+1}g_{r+1}]$. Put $Be_{r+2}g_{r+2} + Be_{r+3}g_{r+3} + \cdots + Be_ng_n = L$, and put $(Be_1g_1 \oplus \cdots \oplus Be_{r-1}g_{r-1} \oplus Q) \cap L = S$, then any element x of S can be written as follows:

$$x = x_1 + x_2 + \cdots + x_{r-1} + x',$$

where $x_i \in Be_ig_i$, $i=1, 2, \ldots, r-1$ and $x' \in Q$. Since Q is injective relative to $N^{\varrho_{r+1}}$ and the annihilator of L in B contains $N^{\varrho_{r+1}}$, a homomorphism $\theta: S \to Q$, defined by putting $\theta(x)=x'$, is extendable to a homomorphism $\Theta: L \to Q$. By Lemma 4.4 Q is a direct summand of M.

For the case Ie let x be an element of $U \cap Be_{r+1}g_{r+1}$, where $U = Be_1g_1 \oplus Be_2g_2 \oplus \cdots \oplus Be_rg_r$. Then we have a representation of x as follows: $x = x_1 + x_2 + \cdots + x_r$, $x_i \in Be_ig_i$. Since $\overline{B}\,\overline{e}_\lambda$ belong to the first kind, every homomorphism $\theta_i: U \cap Be_{r+1}g_{r+1} \to Be_ig_i$, defined by putting $\theta_i(x)=x_i$ for all $x \in U \cap Be_{r+1}g_{r+1}$, is extendable to a homomorphism: $Be_{r+1}g_{r+1} \to Be_ig_i$, $i=1, 2, \ldots, r$. Hence by Lemma 4.4 we have $U + Be_{r+1}g_{r+1} = U \oplus B_{r+1}eg'_{r+1}$, where $U = \sum_{i=1}^{r} Be_ig_i$ and $[Be_{r+1}g_{r+1}] < [Be_{r+1}g_{r+1}]$. Likely as in Ic we can take g'_{r+1} in place of g_{r+1} as a member of a system of generators of M and we shall line up the system of generators of M. If we have again the case Ic for this new system of generators of M, apply the above process to this new system. Nevertheless if we always meet the case Ie, it follows clearly that M is decomposable. Therefore the case left to be proved is following:

$$\begin{aligned} U &= Be_1g_1 + \cdots + Be_rg_r + Be_{r+1}g_{r+1} + \cdots + Be_sg_s \\ &= Be_1g_1 \oplus \cdots \oplus Be_rg_r \oplus Be_{r+1}g_{r+1} \oplus \cdots \oplus Be_sg_s, \end{aligned}$$

and $U \cap Be_{s+1}g_{s+1} \neq 0$, where the S^2-classes of Be_ig_i, $1 \leqq i \leqq r$ are identical, the S^2-classes of Be_jg_j, $r+1 \leqq j \leqq s$ are identical but the S^2-classes of Be_ig_i and Be_jg_j are different each other. According to Proposition 4.8 we have a decomposition of $U + Be_{s+1}g_{s+1}: U + Be_{s+1}g_{s+1} = U' \oplus Q$, where Q is injective relative to $N^{\varrho_{s+1}}$ and $\varrho_{s+1} = [Be_{s+1}g_{s+1}]$. Since $[Be_j] \leqq \varrho_{s+1}$ for $j \geqq s+1$, Q remains to be a direct summand of M. Thus we complete the proof for the case where $\overline{B}\,\overline{e}_\lambda$ is of the first kind.

II. $\overline{B}\,\overline{e}_\lambda$ is the second kind. Let us denote by

$$e_1g_1, \ e_2g_2, \ \ldots, \ e_ng_n \tag{39}$$

the normal sequence of a system of generators of M. Let us consider the case where $Be_1g_1 \cap Be_2g_2 \neq 0$. Let η be the identity isomorphism of $Be_1g_1 \cap Be_2g_2$, i.e. $\eta(x_2)=x_1$, where $x_1=x_2$, $x_1 \in Be_1g_1$ and $x_2 \in Be_2g_2$. If η is maximal, it follows by Proposition 4.9 that $Be_1g_1 + Be_2g_2$ is injective relative to N^{ϱ_2}, where $\varrho_2 = [Be_2g_2]$. Since (39) is a normal sequence of generators, M is decomposable. If η is extendable, likely as in the case Ic, we have a decomposition of $Be_1g_1 + Be_2g_2$ as follows: $Be_1g_1 + Be_2g_2 = Be_1g_1 \oplus Be_2g'_2$, where $[Be_2g'_2] \leqq [Be_1g_1]$.

Next consider the case where $U = Be_1g_1 + Be_2g_2 + \cdots + Be_rg_r = Be_1g_1 \oplus Be_2g_2 \oplus \cdots \oplus Be_rg_r$ and $U \cap Be_{r+1}g_{r+1} = S^1(Be_{r+1}g_{r+1})$, then we can apply Proposition 4.10 to $U + Be_{r+1}g_{r+1}$. If Q is an indecomposable direct summand of $U + Be_{r+1}g_{r+1}$ which is injective relative to $N^{\varrho_{r+1}}$, where $\varrho_{r+1} = [Be_{r+1}g_{r+1}]$, then Q remains as a direct summand of M (cf. Lemma 4.4). If $U \cap Be_{r+1}g_{r+1} = S^2(Be_{r+1}g_{r+1}) \neq 0$, by the remark of Corollary 4.6 the proof for this case is

reduced to the former case. Consequently we finally conclude that M is decomposable. Thus we complete the proof (cf. proof of Proposition 5.1).

THEOREM 5.3. *Let B be a ring with the minimum condition on left ideals. In order that B may be a ring possessing a property such that every finitely generated indecomposable left B-module has a unique minimal submodule it is necessary and sufficient that the following conditions are satisfied:*

I. *B is a left generalized uni-serial ring, i.e. every primitive left ideal has a unique composition series.*

II. *Every isomorphism or its inverse between socles of two quasi-primitive left B-modules is either maximal or extendable to a monomorphism.*

III. *There exist at most two S^2-classes of quasi-primitive left B-modules whose S^1-parts are isomorphic to a given simple left B-module.*

IV. *The number of composition modules of $e_\varkappa N/e_\varkappa N^2$ which are isomorphic to $e_\lambda B/e_\lambda N$ is not larger than two, and if the number is equal to two, there exists only a S^2-class of quasi-primitive left B-modules such that its S^1-part is isomorphic to $Be_\varkappa/Ne_\varkappa$, where e_λ and $e_\varkappa$ are primitive idempotents of B.*

THEOREM 5.4. *Assume that a ring B possesses the same property as in Theorem 5.3. Let $\overline{B}\,\overline{e}_\lambda$ be a simple left B-module. Then there exists an indecomposable B-module which is not quasi-primitive and has $\overline{B}\,\overline{e}_\lambda$ as its S^1-part, if and only if $\overline{B}\,\overline{e}_\lambda$ satisfies one of the following conditions* (i) *and* (ii)*;*

(i) *There exist two S^2-classes of quasi-primitive modules which have $\overline{B}\,\overline{e}_\lambda$ as their S^1-part, i.e. there exist two primitive idempotents $e_\varkappa$, e_μ of B such that $e_\lambda N e_\varkappa \not\equiv 0 \pmod{N^2}$, $e_\lambda N e_\mu \not\equiv 0 \pmod{N^2}$ and $e_\varkappa \neq e_\mu$.*

(ii) *$\overline{B}\,\overline{e}_\lambda$ is of the second kind, i.e. $e_\lambda N e_\varkappa \not\equiv 0 \pmod{N^2}$ and*

$$(e_\lambda N e_\varkappa / e_\lambda N^2 e_\varkappa : e_\varkappa B e_\varkappa / e_\varkappa N e_\varkappa)_r = 2.$$

Now in view of § 2 we have a characterization of SRCRT.

THEOREM 5.5. *Let A be a ring satisfying the minimum condition on right ideals. Let Q be an injective and fully faithful right A-module whose S^1-part is finitely generated. Let us denote by B the inverse A-homomorphism ring of Q. Then A is of SRCRT if and only if B satisfies the all conditions* (I)–(IV) *of Theorem 5.3.*

Let A be a generalized uni-serial ring and N the radical of A. If $e_\lambda N^\varrho e_\varkappa \not\equiv 0$ modula $N^{\varrho+1}$, where $e_\lambda, e_\varkappa$ are primitive idempotents of A. Then from the fact A is left generalized uni-serial it holds that

(i) $\left((e_\lambda N^\varrho e_\varkappa / e_\lambda N^{\varrho+1} e_\varkappa) : (e_\lambda A e_\lambda / e_\lambda N e_\lambda)\right)_l = 1$.

(ii) The map $\varphi : \varkappa \to \lambda$ is univalent.

Similary from the fact A is right generalized uni-serial it holds that

(i)′ $\left((e_\lambda N^\varrho e_\varkappa / e_\lambda N^{\varrho+1} e_\varkappa) : (e_\varkappa A e_\varkappa / e_\varkappa N e_\varkappa)\right)_r = 1$.

(ii)′ The map $\psi : \lambda \to \varkappa$ is univalent.

In case $\varrho = 1$, (ii)′ implies that the S^1-class and the S^2-class of a quasi-primitive left B-module are the same as sets of quasi-primitive left B-modules. (i) and (i)′ imply that every automorphism of $N^{\varrho}e_{\varkappa}/N^{\varrho+1}e_{\varkappa}$ has an extension which is an automorphism of $Ae_{\varkappa}/N^{\varrho+1}e_{\varkappa}$. Then it is easily seen that

(a) by (i), (i)′ and (ii)′ A satisfies the condition II of Theorem 5.3.

(b) by (ii)′ A satisfies the condition III of Theorem 5.3.

(c) by (i)′ A satisfies the condition IV of Theorem 5.3.

Hence every indecomposable left A-module has a unique minimal submodule. Further the conditions (i)′ and (ii)′ shows that there exist no simple left A-module as stated in Theorem 5.4. Therefore every indecomposable left A-module is quasi-primitive. Similary we can prove that every indecomposable right A-module is quasi-primitive. Thus we have the following corollary which was first proved by NAKAYAMA.

COROLLARY 5.6. *Let A be a ring with the minimum condition on left and right ideals. Every indecomposable left A-module as well as every indecomposable right A-module is quasi-primitive if and only if A is a generalized uni-serial.*

Proof. The proof of "if" part has been just explained and "only if" part is a direct consequence of Proposition 2.2.

THEOREM 5.7. *Under the same assumptions and notations of Theorem 5.5 A is a generalized uni-serial ring if and only if B is a generalized uni-serial ring.*

Proof. Let A be a generalized uni-serial ring. Then by Corollary 5.6 we have two conclusions as follows:

(i) A is of SRCRT.

(ii) Every indecomposable right A-module has a unique minimal submodule. Since the duality of modules holds between left B-modules and right A-modules, we have from (i) that every indecomposable left B-module has a unique minimal submodule and B is a left generalized uni-serial ring. On the other hand it follows from (ii) that every indecomposable left B-module is quasi-primitive. Thus, applying Theorem 5.4, we have that B is a right generalized uni-serial ring.

References

[1] KOETHE, G.: Verallgemeinerte Abelsche Gruppen mit Hyperkomplexen Operatorring. Math. Z. **39**, 31–44 (1935). — [2] MORITA, K., Y. KAWADA and H. TACHIKAWA: On injective modules. Math. Z. **68**, 217–226 (1957). — [3] MORITA, K.: Duality for modules and its applications to the theory of rings with minimum condition. Sci. Rep. Tokyo Kyoiku Daigaku **6**, no. 150, 83–142 (1958). — [4] NAKAYAMA, T.: On Frobeniusean algebra II. Ann. of Math. **42**, 1–21 (1941). — [5] NAKAYAMA, T.: Note on uni-serial and generalized uni-serial rings. Proc. Imp. Acad. **16**, 285–289 (1940). — [6] SNNAPPER, S.: Completely indecomposable modules. Cann. J. of Math. **1**, 125–152 (1949). — [7] TACHIKAWA, H.: Duality theorem of character modules for rings with minimum condition. Math. Z. **68**, 479–487 (1958).

Dept. of Math., Nara Gakugei University, Nara (Japan)

(Eingegangen am 29. September 1958)

GÜNZLER, H.
Math. Zeitschr. 71, 223–250 (1959)

Fastperiodische Lösungen linearer hyperbolischer Differentialgleichungen

Von
HANS GÜNZLER

Einleitung

Die von H. BOHR geschaffene Theorie analytischer fastperiodischer Funktionen liefert als Hauptresultat die klassischen Sätze über die Entwicklung derartiger Funktionen in Dirichlet-Reihen. Unter Vermeidung des Wortes „analytisch" lassen sich diese fastperiodischen Funktionen folgendermaßen charakterisieren: $F(z)=F(x+iy)=F(x, y)$ ist eine in einem Streifen $c<x<d$, y beliebig reell, definierte, komplexwertige Funktion, die dort, gleichgradig für alle $x\in(c, d)$, fastperiodisch in y ist und die dem Differentialgleichungssystem

$$\frac{\partial}{\partial x}\operatorname{Re}(F)=\frac{\partial}{\partial y}\operatorname{Im}(F), \qquad \frac{\partial}{\partial y}\operatorname{Re}(F)=-\frac{\partial}{\partial x}\operatorname{Im}(F)$$

genügt, für die also auch die Potentialgleichung $\Delta F=0$ gilt.

Bei dieser Auffassung erhebt sich nun die Frage, ob und mit welchen Ergebnissen eine Theorie fastperiodischer Funktionen $F(x, y)$ durchführbar ist, wenn man statt der Potentialgleichung andere partielle Differentialgleichungen zweiter Ordnung, z. B. solche vom hyperbolischen Typ, zugrunde legt.

Die Resultate und Entwicklungssätze, die man dabei erhalten wird, können natürlich nicht derartig befriedigend und abgerundet wie im Bohrschen Falle sein, da den Lösungen hyperbolischer Differentialgleichungen der Analytizitätscharakter abgeht, dafür ist die behandelbare Funktionenmenge umfassender — bei geeignetem Vorgehen werden auch nicht-differenzierbare Funktionen noch erfaßt — außerdem kann man die so erhaltenen Sätze auf Funktionen einer reellen Veränderlichen anwenden und neue Klassen fastperiodischer Funktionen damit definieren und untersuchen (dies wird in einer demnächst erscheinenden Arbeit geschehen).

Auch wenn man nicht diese Anwendungen im Auge hat, sondern Entwicklungen nach Eigenfunktionen für Lösungen hyperbolischer Differentialgleichungen erhalten möchte, die gleichmäßig konvergieren, darf man sich auf fastperiodische Lösungen solcher Differentialgleichungen beschränken, denn die Eigenfunktionen erweisen sich trivialerweise als fastperiodisch, und bei gleichmäßiger Konvergenz überträgt sich diese Eigenschaft auf die Grenzfunktion.

Im ersten Abschnitt der vorliegenden Arbeit wird zunächst der Begriff der verallgemeinerten Lösung einer hyperbolischen Differentialgleichung

eingeführt (Definition 1); daß dieser Begriff sinnvoll ist, zeigen die Sätze 1.1 und 1.2. Die verallgemeinerten Lösungen bilden dann einen Banach-Raum $\mathfrak{E}$ (Definition 2, Satz 1.3), wobei als Norm der Ausdruck $||F||_\varrho := \sup|\varrho(x)\cdot F(x,y)|$ zugrunde gelegt wird; die „Normfunktion" $\varrho(x)$ kann dabei beliebig positiv vorgegeben werden, jedoch existiert zu jeder Differentialgleichung (1) eine universelle Normfunktion ϱ_0, die alle in Betracht kommenden verallgemeinerten Lösungen einfängt (Satz 1.4). In Abschnitt 2 werden fastperiodische Vektoren aus $\mathfrak{E}$ definiert, d.s. verallgemeinerte Lösungen von (1), die — „ϱ-gleichgradig" in x — in y fastperiodisch sind, anschließend wird ein Approximationssatz für solche Funktionen bewiesen (Satz 3.3): Jede fastperiodische verallgemeinerte Lösung von (1) mit endlicher Norm $||F||_\varrho$ läßt sich im Sinne dieser Norm beliebig genau durch endliche Linearkombinationen von elementaren Funktionen (Eigenfunktionen) der Form $h_\omega(x)\cdot e^{i\omega y}$ approximieren*), wobei die Funktionen $h_\omega(x)$ den zu (1) gehörigen gewöhnlichen Differentialgleichungen

$$h''(x) - \big(a(x) - \omega^2\big)\,h(x) = 0$$

genügen (ω reell). In Abschnitt 4 folgen Sätze über die Stetigkeitseigenschaften fastperiodischer Funktionen und eine Charakterisierung der Fastperiodizität mittels Translationszahlen. Abschnitt 5 ist dann den Fourier-Reihen solcher Funktionen gewidmet, während im letzten Abschnitt einzelne Beispiele zu den vorangegangenen Untersuchungen betrachtet werden.

Betrachtungen, welche mit denen unserer Abhandlung verwandt sind, haben schon verschiedene Autoren angestellt, so BOCHNER und VON NEUMANN [3] 1935, DELSARTE [5] 1938, LEVITAN [10] 1949, MARCENKO [14] 1950 und BEREZANSKIJ [1] 1953. Während in der erstgenannten Arbeit — in Fortsetzung von Arbeiten von MUCKENHOUPT [15] und BOCHNER [2] — die Untersuchung der Fastperiodizitätseigenschaften von Lösungen gewisser Operatorgleichungen sowie Entwicklungssätze dafür Hauptziel sind, haben die späteren Abhandlungen von DELSARTE u.a. vornehmlich die Ausdehnung des Begriffs „fastperiodische Funktion auf der Zahlengeraden" zum Gegenstand, die Betrachtung fastperiodischer Lösungen von hyperbolischen Differentialgleichungen ist hier nur Mittel zum Zweck.

In allen diesen Arbeiten waren gewisse, entweder die behandelbare Funktionenklasse oder die betrachteten Differentialgleichungen einschränkende Voraussetzungen erforderlich: In [3] werden zwar hyperbolische Differentialgleichungen, wie wir sie betrachten, miterfaßt, jedoch sind i.w. nur zweimal differenzierbare Funktionen, die, gleichmäßig in y, gewisse Randbedingungen für $x = c$ und $x = d$ erfüllen, als brauchbare Lösungen zugelassen, ferner ist in dem beweisbaren Approximationssatz nur Konvergenz im Mittel erreichbar und außerdem müssen die betrachteten Funktionen über den ganzen Streifen $c < x < d$ nach x integrabel sein (bei unendlich breitem Streifen also im x-Unendlichen verschwinden). Bei DELSARTE hat man gleichmäßige Konvergenz sowie beliebige stetige Funktionen $F(x, y)$, für die „Lösung" definiert werden

*) ω bedeutet bei h_ω hier und i. f. Index, nicht Ableitung.

kann, aber auf Grund der verwendeten Methoden kann nur eine einzige Differentialgleichung behandelt werden (unser Beispiel 4 in § 6). LEVITAN, MARCENKO und BEREZANSKIJ schließlich können die Ergebnisse von DELSARTE auf ganze Klassen von Differentialgleichungen ausdehnen, jedoch müssen bei ihnen die Koeffizienten der Differentialgleichung im Unendlichen — es wird nur der Fall eines unendlichen x-Intervalles betrachtet — hinreichend stark verschwinden.

Wir machen uns in der vorliegenden Abhandlung von diesen Einschränkungen frei, indem wir von der bisherigen konkreten Auffassung der Funktionen zu einer abstrakten übergehen (wie dies übrigens auch schon in [3] geschehen ist), d.h. wir betrachten die einzelnen Funktionen nicht mehr als bestimmte Lösungen einer partiellen Differentialgleichung, für die wir geeignete Integraldarstellungen suchen, sondern fassen die Funktionen als (fastperiodische) Vektoren eines passenden Banach-Raumes auf. Entscheidendes Hilfsmittel wird dabei ein von W. MAAK in [12] bewiesener Approximationssatz für vektorwertige fastperiodische Funktionen sein, der für die Untersuchung verallgemeinerter fastperiodischer Funktionen infolge der geringen Voraussetzungen, unter denen er dort bewiesen werden konnte, von fundamentaler Bedeutung ist.

§ 1. Verallgemeinerte Lösungen von $F_{xx}-F_{yy}=a(x)\cdot F$

Es werden Lösungen der hyperbolischen Differentialgleichung

$$F_{xx}(x,y)-F_{yy}(x,y)=a(x)\cdot F(x,y) \tag{1}$$

betrachtet[1]). Für spätere Untersuchungen müssen diese Lösungen in einem gewissen Sinne einen vollständigen linearen Raum bilden; der übliche Lösungsbegriff erweist sich dabei als zu eng, man transformiert deshalb (1) auf die Form

$$G_{st}(s,t)=\frac{1}{4}a\left(\frac{s+t}{2}\right)G(s,t), \tag{2}$$

verwandelt diese Differentialgleichung in eine Integralgleichung und betrachtet deren Lösungen (dieses Vorgehen scheint dem Problem angemessener zu sein als z.B. die Verwendung von Distributionen).

Folgende Voraussetzungen und Bezeichnungen werden verwendet: Alle vorkommenden Definitionsbereiche sind der reellen Zahlengeraden R bzw. der reellen Ebene R^2 entnommen; die darauf definierten Funktionen sind komplexwertig und endlich, sie werden, soweit nicht ausdrücklich das Gegenteil vermerkt ist, stets als Lebesguesch-meßbar vorausgesetzt. Statt „eine Funktion hat auf jeder kompakten Teilmenge von M eine gewisse Eigenschaft" sagen wir kurz „die Funktion hat diese Eigenschaft lokal".

[1]) Daß hier nur linear-homogene Differentialgleichungen mit Koeffizienten, die lediglich von x abhängen, in Frage kommen, und daß man sich weiter in gewissem Sinne auf Gleichungen der Form (1) beschränken kann, wird in [7] S. 92/93 auseinandergesetzt; dort finden sich auch Hinweise auf Verallgemeinerungsmöglichkeiten.

Sei nun $I=\{c, d\}$ ein beliebiges Intervall aus R; I kann offen (c, d), halboffen $(c. d]$ bzw. $[c, d)$ oder abgeschlossen $[c, d]$ sein, es kann endlich $(-\infty<c<d<\infty)$ oder unendlich ($c=-\infty$ oder $d=\infty$ bzw. beides) sein; stets sei $c<d$.

Sei weiter $a(x)$ eine auf I definierte, endliche, komplexwertige meßbare Funktion, die dort lokal integrabel ist.

I und $a(x)$ sind im folgenden als fest gegeben zu betrachten.

Definition 1: Eine Funktion $F(x, y)$, die auf dem Streifen

$$P=I\times R=\{(x, y): x\in I \;\&\; -\infty<y \text{ reell } <\infty\}$$

definiert, endlich und komplexwertig ist, die dort außerdem zusammen mit $a(x)\cdot F(x, y)$ lokal integrabel ist, heißt *verallgemeinerte Lösung*[2] *von* (1) *zu* x_0, falls es ein $x_0\in I$ sowie Funktionen $U(s)$, $V(t)$ gibt, so daß auf B

$$(3)\quad \begin{cases} G(s,t)=U(s)+V(t)+ \\ \quad +\frac{1}{4}\iint\limits_{b(s,t)} a\left(\frac{s'+t'}{2}\right)G(s',t')\,ds'\,dt', \end{cases}$$

wo

$$G(s,t):=F\left(\frac{s+t}{2},\ \frac{-s+t}{2}\right)$$

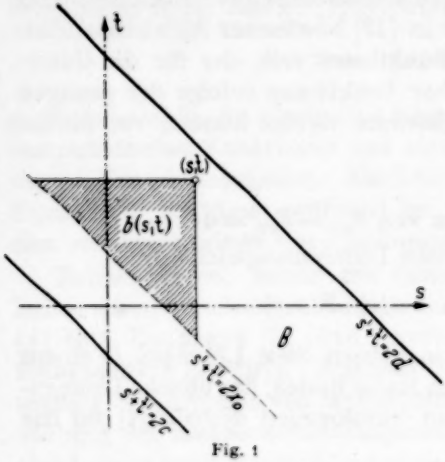


Fig. 1

$U(s), V(t)$ für alle reellen s bzw. t definiert und endlich komplexwertig sind,

B der durch die Drehstreckung

$$2x=s+t,\qquad 2y=-s+t$$

aus P entstehende Streifen ist (vgl. Fig. 1),

$b(s,t)=b(s,t;x_0)$ ein rechtwinkliges abgeschlossenes Dreieck in der s,t-Ebene ist, dessen Katheten achsenparallel sind und im Punkt (s,t) endigen, und dessen Hypotenuse auf der Geraden $s+t=2x_0$ liegt (Fig. 1).

Da $H(s,t):=\iint\limits_{b(s,t)} a\left(\frac{s'+t'}{2}\right)G(s',t')\,ds'\,dt'$ auf B stetig, also sicher lokal integrabel ist, folgt mit FUBINI aus (3), daß sowohl $U(s)$ als auch $V(t)$ lokal integrabel sind.

Nach obiger Definition scheint in der Aussage „die Funktion F ist verallgemeinerte Lösung von (1)" die Gerade $x=x_0$ eine ausgezeichnete Rolle zu spielen. Daß dem nicht so ist, besagt der folgende

[2] Jede verallgemeinerte oder schwache Lösung $F(x, y)$ im Sinne von SOBOLEW ist auch verallgemeinerte Lösung nach Definition 1 hier — jedenfalls soweit es sich um stetige $F(x, y)$ handelt, und die später zu untersuchenden fastperiodischen $F(x, y)$ sind automatisch stetig; das Umgekehrte gilt nicht. — Die wesentlichsten Eigenschaften des hier verwendeten Lösungsbegriffs sind z. B. in [7], Anhang III, zusammengestellt (insbesondere Existenz- und Eindeutigkeitssätze).

Satz 1.1. *Sind x_0, x_1 irgend zwei Zahlen aus I, und ist F verallgemeinerte Lösung von* (1) *zu x_0, so ist F auch verallgemeinerte Lösung von* (1) *zu x_1.*

Der elementare Beweis möge hier übergangen werden, er findet sich in [7], S. 13.

Man kann daher im folgenden den Zusatz „zu x_0" weglassen und kurz von *verallgemeinerten Lösungen von* (1) sprechen.

Als nächstes soll gezeigt werden, daß Definition 1 wirklich eine Verallgemeinerung des üblichen Lösungsbegriffes liefert:

Satz 1.2. *Ist $F(x, y)$ auf P zweimal stetig differenzierbar und dort Lösung von* (1), *so ist dort F verallgemeinerte Lösung von* (1). *Ist umgekehrt F auf P verallgemeinerte Lösung von* (1), *ist weiter F in einer vollen Umgebung*[3]) *von $(x_1, y_1) \in P$ zweimal stetig differenzierbar, und ist $a(x)$ in x_1 stetig*[4]), *so genügt F in (x_1, y_1) der Differentialgleichung* (1).

Beweis. Ist $F(x, y)$ auf P zweimal stetig differenzierbar, so gilt das entsprechende auch von $G(s, t)$ auf B; unter Beachtung von (1) rechnet man leicht nach, daß dann G der Gleichung

$$G_{ts}(s,t) = \frac{1}{4}\, a\left(\frac{s+t}{2}\right) G(s,t) \tag{4}$$

genügt.

Ist nun x_0 beliebig aus I gewählt, so gilt nach FUBINI (vgl. Fig. 1)

$$\iint\limits_{b(s,t)} \ldots ds'\, dt' = \int\limits_{2x_0-s}^{t} dt' \int\limits_{2x_0-t'}^{s} \ldots ds' = \int\limits_{2x_0-t}^{s} ds' \int\limits_{2x_0-s'}^{t} \ldots dt', \tag{5}$$

also erhält man aus (4)

$$\begin{aligned}
\frac{1}{4} \iint\limits_{b(s,t)} a\left(\frac{s'+t'}{2}\right) G(s',t')\, ds'\, dt' &= \frac{1}{4} \int\limits_{2x_0-s}^{t} \int\limits_{2x_0-t'}^{s} G_{ts}(s',t')\, ds'\, dt' \\
&= \int\limits_{2x_0-s}^{t} \big(G_t(s,t') - G\,(2x_0 - t', t')\big)\, dt' \\
&= G(s,t) - G(s, 2x_0 - s) - \int\limits_{2x_0-s}^{t} G_t(2x_0 - t', t')\, dt' \\
&= -\,U(s) - V(t) + G(s,t),
\end{aligned}$$

wenn man z. B.

$$\begin{aligned}
U(s) &:= G(s, 2x_0 - s) + \int\limits_{2x_0-s}^{0} G_t(2x_0 - t', t')\, dt', \\
V(t) &:= \int\limits_{0}^{t} G_t(2x_0 - t', t')\, dt'
\end{aligned}$$

setzt; das zu F gehörige G genügt also (3) auf B (alle übrigen Voraussetzungen über G sind wegen der Stetigkeit von F und G trivialerweise erfüllt), F ist damit verallgemeinerte Lösung von (1).

[3]) Wie Satz und Beweis abzuändern sind, wenn (x_1, y_1) auf dem Rand von P liegt, ist klar.

[4]) Wegen (1) ist, falls $F(x_1, y_1) \neq 0$, diese Voraussetzung im allgemeinen eine notwendige.

Sei nun umgekehrt F verallgemeinerte Lösung von (1), d.h. für das zugehörige $G(s,t)$ gilt mit passenden $U(s)$ und $V(t)$

$$(3') \qquad G(s,t) = U(s) + V(t) + \frac{1}{4}\iint\limits_{b(s,t)} a\,G\,ds'\,dt' =: U(s) + V(t) + \frac{1}{4}H(s,t).$$

Wir zeigen nun zunächst den auch für später wichtigen

Hilfssatz. *Ist $G(s,t)$ im Punkt (s_1,t_1) stetig*[5]*, ist $G(s,t)$ auf einer offenen Menge M, die das Dreieck $b(s_1,t_1)$ enthält, beschränkt, sonst auf B zusammen mit $a\cdot G$ lokal integrabel, so ist das gemäß (3') G zugeordnete $H(s,t)$ im Punkt (s_1,t_1) nach s und t differenzierbar und es gelten die Formeln (6) und (6')*[3].

Zum Beweis wählen wir ein Dreieck $b_2 = b(s_2,t_2)$ aus M, so daß $b(s_1,t_1)$ — mit Ausnahme seiner Hypotenuse — ganz im Innern von b_2 liegt[3]; in b_2 wird sich alles folgende abspielen. Da $H(s,t)$ überall stetig ist, sind $U(s)$ und $V(t)$ auf b_2 beschränkt, so daß also alle folgenden Integrale für (s,t) aus b_2 existieren

$$\begin{aligned} H(s,t) &= \iint\limits_{b(s,t)} a\left(\frac{s'+t'}{2}\right) G(s',t')\,ds'\,dt' \\ &= \iint\limits_{b(s,t)} a\left(\frac{s'+t'}{2}\right) U(s')\,ds'\,dt' + \iint\limits_{b(s,t)} a\,V(t')\,ds'\,dt' + \frac{1}{4}\iint\limits_{b(s,t)} a\,H(s',t')\,ds'\,dt' \\ &=: I + II + III. \end{aligned}$$

Für I erhält man mit (5)

$$I = \int\limits_{2x_0-s}^{t} W(s,t')\,dt', \qquad W(s,t') := \int\limits_{2x_0-t'}^{s} a\left(\frac{s'+t'}{2}\right) U(s')\,ds'.$$

Der Integrand $W(s,t')$ ist dabei in t' für $t' = t_1$ stetig (s fest $= s_1$), wie die folgenden Abschätzungen zeigen:

$$\begin{aligned} |W(s_1,t') - W(s_1,t_1)| &= \left| \int\limits_{2x_0-t'}^{s_1} a\left(\frac{s'+t'}{2}\right) U(s')\,ds' - \int\limits_{2x_0-t'}^{s_1} a\left(\frac{s'+t_1}{2}\right) U(s')\,ds' + \right. \\ &\qquad \left. + \int\limits_{2x_0-t'}^{s_1} a\left(\frac{s'+t_1}{2}\right) U(s')\,ds' - \int\limits_{2x_0-t_1}^{s_1} a\left(\frac{s'+t_1}{2}\right) U(s')\,ds' \right| \\ &\leqq \left| \int\limits_{2x_0-t'}^{s_1} \left(a\left(\frac{s'+t'}{2}\right) - a\left(\frac{s'+t_1}{2}\right)\right) U(s')\,ds' \right| + \left| \int\limits_{2x_0-t_1}^{2x_0-t'} a\left(\frac{s'+t_1}{2}\right) U(s')\,ds' \right| \\ &\leqq \sup_{(s,t)\in b_2} |U(s)| \cdot \left(\left| \int\limits_{2x_0-t_2}^{s_1} \left| a\left(\frac{s'+t'}{2}\right) - a\left(\frac{s'+t_1}{2}\right) \right| ds' \right| + \left| \int\limits_{2x_0-t_1}^{2x_0-t'} \left| a\left(\frac{s'+t_1}{2}\right) \right| ds' \right| \right), \end{aligned}$$

[5]) Nach (3) genügen s-Stetigkeit in $s = s_1$ (t fest $= t_1$) und t-Stetigkeit in $t = t_1$ (s fest $= s_1$).

beide in der letzten Klammer vorkommenden Integrale gehen für $t' \to t_1$ gegen 0 nach bekannten Sätzen der Maßtheorie (vgl. [7], S. 75). Somit ist I in (s_1, t_1) nach t differenzierbar und hat als Ableitung $W(s_1, t_1)$.

Weiter wird nach (5)

$$II = \int_{2x_0 - s}^{t} V(t') \int_{2x_0 - t'}^{s} a\left(\frac{s'+t'}{2}\right) ds'\, dt' = \int_{2x_0 - s}^{t} V(t') \int_{2x_0}^{s+t'} a\left(\frac{r}{2}\right) dr\, dt'.$$

Da nach Voraussetzung und wegen der Stetigkeit von H auch die U und V in s_1 bzw. t_1 stetig sind, ist der Integrand hier ebenfalls in $t' = t_1$ stetig, man darf also auch hier nach den üblichen Regeln differenzieren.

Schließlich ist

$$III = \int_{2x_0 - s}^{t} \int_{2x_0 - t'}^{s} a\left(\frac{s'+t'}{2}\right) \cdot H(s', t')\, ds'\, dt';$$

unter Benutzung der gleichmäßigen Stetigkeit von H auf b_2 zeigt man hier durch analoge Abschätzungen wie im Falle I die Stetigkeit des Integranden von $\int \ldots dt'$ in $t' = t_1$, III ist dort auch nach t differenzierbar.

Durch Zusammenfassung der für die drei Ableitungen von I, II, III erhaltenen Ausdrücke bekommt man dann

$$H_t(s_1, t_1) = \int_{2x_0 - t_1}^{s_1} a\left(\frac{s'+t_1}{2}\right) G(s', t_1)\, ds'. \tag{6}$$

Genauso zeigt man die Existenz von $H_s(s_1, t_1)$ und die Gültigkeit einer entsprechenden Gleichung

$$H_s(s_1, t_1) = \int_{2x_0 - s_1}^{t_1} a\left(\frac{s_1+t'}{2}\right) G(s_1, t')\, dt'. \tag{6'}$$

Damit können wir jetzt den Beweis von Satz 1.2 vollenden: Da F um (x_1, y_1) jedenfalls stetig ist, ist es dort beschränkt, dasselbe gilt von dem zugehörigen G um das entsprechende (s_1, t_1); wählen wir also $x_0 = \frac{s_1+t_1}{2}$, was nach Satz 1.1 möglich ist, so sind die Voraussetzungen unseres Hilfssatzes in einer ganzen Umgebung von (s_1, t_1) erfüllt, so daß in einer ganzen derartigen Umgebung (6) und (6') gelten.

Da weiter nach Voraussetzung $a(x)$ in $x = x_1 = \frac{s_1+t_1}{2}$ stetig ist, zeigt (6), daß auch noch

$$H_{ts}(s_1, t_1) \text{ existiert } = a\left(\frac{s_1+t_1}{2}\right) G(s_1, t_1).$$

Nun ist auf Grund der über F gemachten Voraussetzungen G um (s_1, t_1) differenzierbar; wegen $G = U + V + \frac{1}{4}H$ ist daher V um t_1 differenzierbar und man hat um (s_1, t_1)

$$G_t(s, t) = V'(t) + \frac{1}{4} H_t(s, t), \qquad G_{ts}(s_1, t_1) = \frac{1}{4} H_{ts}(s_1, t_1) = \frac{1}{4} a\left(\frac{s_1+t_1}{2}\right) G(s_1, t_1).$$

Daraus folgt aber durch eine einfache Umrechnung wieder Gl. (1) im Punkte (x_1, y_1).

Als nächstes sollen nun gewisse Teilmengen von verallgemeinerten Lösungen von (1) zu einem normierten Raum $\mathfrak{E}$ zusammengefaßt werden:

Sei $\varrho(x)$ eine fest vorgegebene reellwertige Funktion, die auf I definiert ist, deren Werte lokal, d.h. auf jedem kompakten Teilintervall I_1 von I, größer als eine positive Zahl δ_{I_1} sind, die aber nicht meßbar oder endlich zu sein braucht (*„Normfunktion"*); damit werde festgesetzt:

Definition 2.

$$||F(x,y)||_\varrho := ||F||_\varrho := \sup_{(x,y)\in P} \varrho(x)\cdot|F(x,y)| \qquad \textit{(Norm von } F\textit{)}$$

$$\mathfrak{E} := \mathfrak{E}(I, a, \varrho) := \{F(x,y): F \text{ ist verallg. Lösung von (1) } \&\ ||F||_\varrho < \infty\}.$$

Für F aus $\mathfrak{E}$ und reelles c

$$T_c F(x,y) := F(x, y+c).$$

Es gilt nun

Satz 1.3. *$\mathfrak{E}$ ist ein Banach-Raum, d.h. ein normierter, vollständiger linearer Vektorraum über dem Körper der komplexen Zahlen.*

Die T_c sind lineare Operatoren, die die Norm invariant lassen und die $\mathfrak{E}$ auf sich abbilden. Sie bilden eine der additiven Gruppe der reellen Zahlen isomorphe Gruppe $\mathfrak{G} := (T_c: c \text{ reell})$, $T_c T_d = T_{c+d}$.

Beweis. Daß $\mathfrak{E}$ ein normierter Vektorraum ist, ist trivial nachzuprüfen (Gleichheit in $\mathfrak{E}$ bedeutet Gleichheit der betreffenden Funktionen in allen Punkten des Streifens P). Eine einfache Substitution in (3) zeigt, daß $T_c F$ wieder einer Gleichung der Form (3) genügt, d.h. $\mathfrak{E}$ ist invariant gegenüber $\mathfrak{G}$; die übrigen behaupteten Eigenschaften von $\mathfrak{E}$ sind ebenfalls klar, so daß lediglich die Vollständigkeit von $\mathfrak{E}$ nachzuweisen ist:

Sei also (F_n) eine Fundamentalfolge aus $\mathfrak{E}$; auf Grund der Eigenschaften von ϱ sind die F_n auf jedem Teilstreifen $I_1 \times R$ (I_1 kompaktes Teilintervall von I) beschränkt und konvergieren dort gleichmäßig gegen eine Grenzfunktion F_∞, die dort dann ebenfalls beschränkt und meßbar ist; F_∞ ist daher ebenso wie $a \cdot F_\infty$ lokal integrierbar.

Dieses F_∞ ist auch im Sinne der Norm Grenzwert der F_n, denn ist N so gewählt, daß für n, $m > N$ stets $||F_n - F_m||_\varrho < \varepsilon$, und, bei beliebig vorgegebenem $(x,y) \in P$, $n = n(x,y) > N$ so, daß

$$\varrho(x)\cdot|F_\infty(x,y) - F_n(x,y)| < \varepsilon,$$

so hat man für $m > N$

$$\varrho(x)\cdot|F_\infty(x,y) - F_m(x,y)| \leqq \varrho(x)\cdot\left(|F_\infty(x,y) - F_n(x,y)| + |F_n(x,y) - F_m(x,y)|\right)$$
$$\leqq \varepsilon + ||F_m - F_n||_\varrho \leqq 2\varepsilon.$$

Es ist jetzt nur noch zu zeigen, daß das F_∞ entsprechende G_∞ einer Gleichung der Form (3) genügt. Zunächst konvergieren die $G_n(s,t)$ lokal auf B

gleichmäßig gegen $G_\infty(s,t)$; eine einfache Abschätzung zeigt, daß dann dasselbe für die Funktionen $H_n(s,t) := \iint\limits_{b(s,t)} a \cdot G_n \, ds' \, dt'$ und $H_\infty(s,t)$ gilt, so daß also auch die Funktionen $W_n(s,t) := U_n(s) + V_n(t)$ lokal gleichmäßig gegen W_∞ konvergieren. W_∞ ist von der Form $U(s)+V(t)$: Es konvergiert nämlich $W_n(s,t_0) := u_n(s)$ (t_0 beliebig, fest) gleichmäßig in einem s-Intervall der Art, wie es in Fig. 2 angedeutet ist. Geht man daher von den Funktionen U_n, V_n über zu $u_n(s)$, $v_n(t) := V_n(t) - V_n(t_0)$, so ist wieder $W_n = u_n + v_n$; da W_n auch auf dem zu s_0 gehörigen t-Intervall (vgl. Fig. 2) konvergiert und $u_n(s_0)$ nach Konstruktion konvergiert, hat man jetzt erreicht, daß in diesem t-Intervall $v_n(t)$ ebenfalls gleichmäßig konvergiert; durch Wahl von neuen Punkten $t_1, s_1, t_2, \ldots$ kann man so einsehen, daß sowohl die Funktionen $u_n(s)$ als auch die Funktionen $v_n(t)$ überall lokal gleichmäßig konvergieren, so daß also auch $W_\infty(s,t)$ von der Form $u_\infty(s)+v_\infty(t)$ ist. q.e.d.

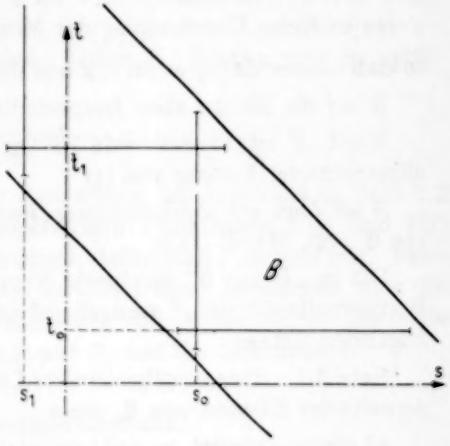


Fig. 2

$\mathfrak{E}$ ist übrigens auch noch gegenüber der Spiegelung S, $(SF)(x,y) := F(x,-y)$ invariant, also auch gegenüber der von $\mathfrak{G}$ und S erzeugten, nichtabelschen Gruppe $\mathfrak{G}_S = \mathfrak{G} \cup S\mathfrak{G}$, deren Schranke ebenfalls 1 ist.

Bei der Untersuchung der Räume $\mathfrak{E}$ muß man nun nicht beliebige Normfunktionen $\varrho(x)$ zulassen, sondern man kann sich hier beschränken:

Satz 1.4. *Zu jedem lokal integrablem $a(x)$ existiert eine „universelle" Normfunktion $\varrho_0(x)$, gegeben durch*

$$(7) \qquad \varrho_0(x) := \frac{1}{1+|x|}\, e^{-x-x_0 \cdot \psi(x)}, \quad wo \quad \psi(x) := \int\limits_{x_0}^{x} |a(\xi)|\, d\xi, \quad x_0 \text{ bel. fest} \in I,$$

so daß folgendes gilt:

Ist $F(x,y)$ verallgemeinerte Lösung von (1) und gibt es zu $F(x,y)$ eine Normfunktion $\varrho(x)$, so daß $\|F\|_\varrho < \infty$, so ist auch schon $\|F\|_{\varrho_0} < \infty$, kurz

$$\bigcup_{\varrho} \mathfrak{E}_\varrho = \mathfrak{E}_{\varrho_0}.$$

Die ϱ_0-Norm ist sogar gleichmäßig durch jede ϱ-Norm beschränkt, d.h. es existiert ein (nur von den Funktionen $a(x)$, $\varrho(x)$ sowie von x_0 abhängige) Konstante C, so daß für alle verallgemeinerten Lösungen von (1) gilt

$$\|F\|_{\varrho_0} \leqq C \cdot \|F\|_\varrho.$$

Aus der Tatsache, daß eine Lösung von (1) überhaupt in y beschränkt ist für jedes feste x, folgt also, daß diese Schranken in ihrer Abhängigkeit von x nicht zu groß werden können, so daß Satz 1.4 eine Aussage über die

Wachstumseigenschaften von Lösungen linearer hyperbolischer Differentialgleichungen macht. [Für $a(x)=0$ ist das oben angegebene $\varrho_0(x)$ übrigens scharf.]

Da wir Satz 1.4 im folgenden nicht benötigen, wollen wir auf seinen Beweis hier nicht weiter eingehen.

§ 2. Fastperiodische Lösungen

Hat man einen Banach-Raum $\mathfrak{E}$ und eine Gruppe $\mathfrak{G}$ linearer Operatoren, die auf $\mathfrak{E}$ wirken, so ist erklärt, wann ein Element F aus $\mathfrak{E}$ fastperiodisch heißt (vgl. z.B. MAAK [12]):

Definition 3. Ein F aus $\mathfrak{E}$ heißt *fastperiodisch*, wenn es zu jedem positiven ε eine endliche Überdeckung der Menge der reellen Zahlen R gibt, $R=\bigcup_{\nu=1}^{m} A_\nu$, so daß immer dann, wenn c, d aus demselben A_ν sind, $\|T_c F - T_d F\|_\varrho < \varepsilon$.

$\mathfrak{F}$ sei die Menge aller fastperiodischen Vektoren von $\mathfrak{E}$.

Statt „F ist Element von $\mathfrak{F}$" sagen wir auch „F ist fastperiodische (verallgemeinerte) Lösung von (1)".

$\mathfrak{F}$ ist dann ein abgeschlossener, linearer, gegenüber $\mathfrak{G}$ invarianter Unterraum von $\mathfrak{E}$ (vgl. MAAK [12]).

Die in „F aus $\mathfrak{E}$" steckende Voraussetzung $\|F\|_\varrho<\infty$ ist nun infolge der Fastperiodizität der F weitgehend von selbst erfüllt; dies ist der Inhalt des folgenden Satzes

Satz 2.1. *Eine verallgemeinerte Lösung $F(x, y)$ von (1) ist genau dann fastperiodisches Element von $\mathfrak{E}$, wenn*

a) ein y_0 existiert, so daß $\sup_{x\in I} |\varrho(x)\, F(x, y_0)|<\infty$ und

b) zu jedem positiven ε eine endliche Überdeckung $\bigcup_{\nu=1}^{m} A_\nu$ von R vorhanden ist, so daß für $c, d\in A_\nu$ stets

$$\|T_c F - T_d F\|_\varrho < \varepsilon.$$

Daß a) und b) notwendig für F aus $\mathfrak{F}$ sind, ist klar; diese Bedingungen sind hinreichend, wenn man gezeigt hat, daß aus a) und b) $\|F\|_\varrho<\infty$ folgt.

Es gilt aber ganz allgemein: *Ist $\Phi_i(y)$ eine Funktionenmenge, die gleichgradig in $i\in J$ fastperiodisch ist, und ist $\Phi_i(y_0)$ für $i\in J$ beschränkt, so ist $\Phi_i(y)$ überhaupt beschränkt.*

Auf die genauere Formulierung dieses Satzes und auf seinen Beweis soll hier verzichtet werden (er findet sich in [7], S. 22), da Satz 2.1 nicht weiter verwendet wird.

Im übrigen kann man Voraussetzung a) hier nicht weglassen, wie einfache Gegenbeispiele zeigen. Mit Hilfe von Satz 1.4 kann man jedoch aus b) allein eventuell schließen, daß $F(x, y)$ fastperiodischer Vektor bezüglich der universellen ϱ_0-Norm ist [denn aus b) folgt, daß $F(x, y)$ bei festem x beschränkt ist]; vgl. hierzu auch Satz 3.4.

Nur eine andere Formulierung der Definition 2 ist (wegen der hier vorkommenden Begriffe vgl. MAAK [11], S. 150)

Satz 2.2. *Ein Vektor F aus $\mathfrak{E}$ ist genau dann fastperiodisch, wenn die Menge $\mathfrak{G}F$ totalbeschränkt oder, was hier gleichwertig, bedingt kompakt*[6]) *in $\mathfrak{E}$ ist.*

Eine Charakterisierung der fastperiodischen Elemente von $\mathfrak{E}$ mittels Translationseigenschaften wird im 4. Abschnitt (Satz 4.2) gegeben werden; auch die Stetigkeitseigenschaften solcher F können erst später (Satz 4.1) mittels des Approximationssatzes abgeleitet werden, hier sei nur soviel bemerkt:

Gemäß der Darstellung (3) von $G(s, t)$ ist $F(x, y) = G(x - y, x + y)$ bei festem x bzw. y stets meßbar; ist F fastperiodischer Vektor, so ist $F(x, y)$ bei festem x im Maakschen Sinne fastperiodische Funktion von y, die meßbar ist; z. B. nach MAAK [11], S. 86, Satz 2 ist $F(x, y)$ dann in y stetig (sogar gleichmäßig stetig).

Weiter stellen wir für später noch fest: Jedem $F \in \mathfrak{E}$ kann man in natürlicher Weise durch die Zuordnung

$$F \to F_g = \tfrac{1}{2}(F + SF) = \tfrac{1}{2}\big(F(x, y) + F(x, -y)\big) \tag{8}$$

ein *gerades* F_g zuordnen, das wegen der Bemerkung im Anschluß an Satz 1.2 ebenfalls in $\mathfrak{E}$ liegt und für das die zugehörigen Funktionen $U(s)$ und $V(t)$ gleich sind (dies ist übrigens charakteristisch dafür, daß F gerade ist). Fastperiodische Vektoren gehen dabei wieder in solche über; bezeichnet man die Menge aller geraden fastperiodischen Elemente von $\mathfrak{F}$ mit $\mathfrak{F}_g$, so gilt: Die Zuordnung (8) ist eine Projektion von $\mathfrak{F}$ auf $\mathfrak{F}_g$ mit der Schranke 1.

§ 3. Der Approximationssatz

Für fastperiodische Vektoren eines Banach-Raumes E mit der Transformationsgruppe G gilt nach MAAK [12], § 1 und § 2 ein Approximationssatz, der besagt, daß der Raum der fastperiodischen Elemente von E gleich der abgeschlossenen Hülle des von allen endlich-dimensionalen, gegenüber G invarianten, irreduziblen Teilräumen aufgespannten linearen Raumes ist.

Wir wollen diesen Satz hier anwenden und bestimmen deshalb zunächst einmal die endlich-dimensionalen, irreduziblen, invarianten Unterräume von $\mathfrak{E}$ bezüglich $\mathfrak{G}$:

Jeder solche Unterraum definiert eine irreduzible Darstellung von $\mathfrak{G}$, die nach dem Schurschen Lemma (MAAK [11], § 5), da $\mathfrak{G}$ abelsch ist, eindimensional ist, d. h. die endlich-dimensionalen invarianten Unterräume von $\mathfrak{E}$ werden jeweils von einem einzigen F aufgespannt, das dann wegen der Invarianz einer Gleichung der Form

$$T_c F(x, y) = k(c) \cdot F(x, y) \qquad T_c \text{ aus } \mathfrak{G} \tag{9}$$

oder

$$F(x, y + c) = k(c) \cdot F(x, y) \qquad \text{für } x \in I;\ y, c \in R$$

[6]) Dies ist die von Neumannsche Definition von fastperiodisch.

genügen muß. Eine verallgemeinerte Lösung von (1), die (9) erfüllt, möge „*elementar*" genannt werden, ihre Bestimmung wird durch Satz 3.1 und 3.2 erledigt:

Satz 3.1. *Eine Funktion $F(x, y)$ ist genau dann elementare verallgemeinerte Lösung von* (1), *wenn sie von der Form*

$$F(x, y) = h_l(x) \cdot e^{ly} \tag{10}$$

ist, wo l beliebig komplex und $h_l(x)$ auf I eine Lösung der folgenden Differentialgleichung

$$h''(x) - h(x) \cdot \left(a(x) + l^2\right) = 0 \tag{11}$$

ist, d.h. $h_l(x)$ ist auf I einmal stetig differenzierbar, $(h_l)'$ ist dort lokal totalstetig, h_l genügt f.ü. (11).

Hierzu ist zu bemerken, daß aus der Totalstetigkeit von h_l' die Existenz von h_l'' f. ü. auf I folgt; aus (11) folgt, daß an Stetigkeitsstellen von $a(x)$ stets h_l'' existiert und durch (11) gegeben ist.

Für stetige $a(x)$ ist ein F also elementar, wenn es von der Form (10) mit zweimal stetig differenzierbarem h ist, das (11) auf I genügt.

Beweis. Das einem F der Form (10) zugeordnete G genügt f.ü. auf B der Gl. (2) $G_{st} = \frac{1}{4} aG$, wie man sofort nachrechnet. Genauso wie beim Beweis des ersten Teils von Satz 1.2 folgert man daraus durch Integration Gl. (3) [die Formel $\int_{s_1}^{s_2} f'(s)\, ds = f(s_2) - f(s_1)$ kann dabei wie dort angewendet werden, da G_t in s mit h_l' lokal totalstetig ist]. Außerdem genügt ein solches F offenbar (9).

Sei nun umgekehrt $F \not\equiv 0$ verallgemeinerte Lösung von (1), die (9) erfüllt; dann ist jedenfalls (vgl. die Bemerkung im Anschluß an Satz 2.2) $k(c)$ meßbar und lokal integrabel, außerdem folgt für $k(c)$ aus (9) $k(c+d) = k(c) \cdot k(d)$. Solche Funktionen haben aber (vgl. MAAK [11], S. 71) die Form e^{lc} mit komplexem l, so daß man $F(x, y+c) = F(x, y) \cdot e^{lc}$ hat. Für $y = 0$ und $l = y$ erhält man daraus

$$F(x, y) = F(x, 0) \cdot e^{ly} =: h_l(x) \cdot e^{ly}$$

mit meßbarem, lokal integrablem $h_l(x)$.

Zur Bestimmung von h_l, wir schreiben dafür im folgenden kurz h, gehen wir in (3) ein: Das zu F gehörige G hat die Form

$$G(s, t) = h\left(\frac{s+t}{2}\right) \cdot \exp\left(l \cdot \left(\frac{-s+t}{2}\right)\right) \qquad \left(\exp(z) := e^z\right), \tag{12}$$

also wird (3) zu

$$h\left(\frac{s+t}{2}\right) \cdot \exp\left(l \cdot \left(\frac{-s+t}{2}\right)\right) = U(s) + V(t) + \frac{1}{4} \iint_{b(s,t)} a\left(\frac{s'+t'}{2}\right) G(s', t')\, ds'\, dt' \tag{13}$$

$$= U + V + \frac{1}{4} H.$$

Zunächst soll nun gezeigt werden, daß die Funktionen U, V, G, h stetig sind.

Sei s_0 beliebig reell; wir wählen dazu ein t-Intervall I_1, $I_1=[t_1, t_2]$ so, daß die Strecke $(s_0)\times I_1$ vom Rand des Streifens B positiven Abstand hat (Fig. 3); integriert man (13) nach t über I_1, so erhält man für s nahe s_0

$$e^{-ls}\int_{I_1}\exp\left(\frac{l}{2}(s+t)\right)\cdot h\left(\frac{s+t}{2}\right)dt=(t_2-t_1)\,U(s)+\int_{I_1}V(t)\,dt+\frac{1}{4}\int_{I_1}H(s,t)\,dt$$

oder, wenn man links $s+t=:\tau$ setzt,

$$(14)\quad\begin{cases}e^{-ls}\displaystyle\int_{t_1+s}^{t_2+s}e^{l\tau/2}h\left(\frac{\tau}{2}\right)d\tau=(t_2-t_1)\,U(s)+\\ \qquad+\displaystyle\int_{I_1}V(t)+\frac{1}{4}\int_{I_1}H(s,t)\,dt.\end{cases}$$

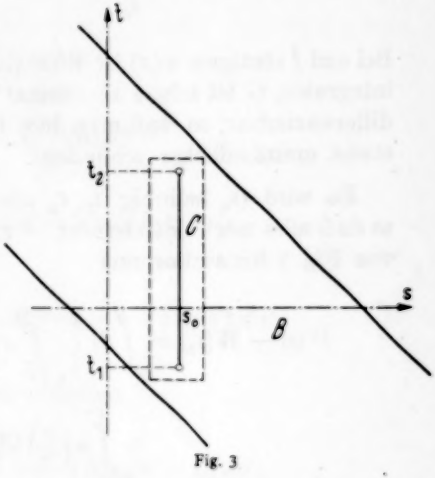


Fig. 3

$H(s,t)$ ist auf B stetig, also ist das letzte Glied rechts bei s_0 stetig; die linke Seite ist ebenfalls bei s_0 stetig, daher ist nach (14) auch $U(s)$ in s_0 stetig; s_0 war beliebig, somit ist $U(s)$ überall stetig. Genauso erschließt man die Stetigkeit von $V(t)$ und damit auch die von $G(s,t)$ auf ihren Definitionsbereichen. Nach (12) ist damit aber auch $h(x)$ auf I stetig.

Die Funktionen U, V, G und h sind auf ihren Definitionsbereichen einmal stetig differenzierbar

Das Argument des Integrals auf der linken Seite von (14) ist stetig, also ist das Integral und damit die ganze linke Seite von (14) um s_0 nach s differenzierbar (s_0 und I_1 wieder wie oben gewählt); aus der Stetigkeit von $G(s,t)$ folgt weiter, daß $G(s,t)$ lokal beschränkt ist, so daß der beim Beweis von Satz 1.2 formulierte Hilfssatz überall auf B anwendbar ist: $H(s,t)$ ist überall auf B nach s differenzierbar, nach Gl. (6') ist H_s auch noch stetig[7]). Nach bekannten Sätzen über die Differentiation von Parameterintegralen ist dann das rechte Integral von (14) ebenfalls um s_0 nach s stetig differenzierbar; also ist nach (14) auch $U(s)$ um s_0 stetig differenzierbar. Weil s_0 beliebig war, ist $U(s)$ und genauso auch $V(t)$ überall auf R stetig differenzierbar, mit (3) und (12) sind dann auch die Funktionen $G(s,t)$ und $h(x)$ auf B bzw. I einmal stetig differenzierbar.

Durch Differentiation nach s erhält man aus (14) um s_0

$$(15)\quad\begin{cases}-l\cdot e^{-ls}\displaystyle\int_{t_1+s}^{t_2+s}e^{l\tau/2}\,h\left(\frac{\tau}{2}\right)d\tau+e^{-ls}\left[e^{l\tau/2}h\left(\frac{\tau}{2}\right)\right]_{\tau=t_1+s}^{\tau=t_2+s}\\ \qquad=(t_2-t_1)\,U'(s)+\dfrac{1}{4}\displaystyle\int_{I_1}H_s(s,t)\,dt.\end{cases}$$

[7]) Um dies zu erkennen, führe man in dem dortigen Integral $r:=s+t'$ als neue Integrationsveränderliche ein, wegen der gleichmäßigen Stetigkeit von G kann man dann leicht die st-Stetigkeit von H_s einsehen.

Um nachzuweisen, daß U und V zweimal stetig (!) differenzierbar sind, hat man nur das rechte Integral in (15) zu betrachten, das nach (6') — man führt wieder $\tau := s + t'$ als neue Integrationsveränderliche ein — die Form hat

$$W(s) := \int_{I_1} H_s(s,t)\,dt = \int_{t_1}^{t_2}\int_{2x_0}^{s+t} a\left(\frac{\tau}{2}\right) G(s,\tau - s)\,d\tau\,dt.$$

Bei auf I stetigem $a(x)$ ist $W(s)$ (Satz über die Differentiation von Parameterintegralen, G ist schon als einmal stetig differenzierbar erkannt) sicher stetig differenzierbar, so daß man hier fertig wäre. Im allgemeinen Fall muß man etwas umständlicher schließen:

Es wird (s_0 beliebig, t_1, t_2 wie oben gewählt, s hinreichend nahe bei s_0, so daß alles noch definiert ist — z. B. so, daß man nicht über das Rechteck C von Fig. 3 hinauskommt)

$$\begin{aligned} W(s) - W(s_0) = \int_{t_1}^{t_2} dt \Bigg\{ & \int_{s_0+t}^{s+t} a\left(\frac{\tau}{2}\right)\cdot\big(G(s,\tau - s) - G(s_0,\tau - s_0)\big)\,d\tau + \\ & + \int_{s_0+t}^{s+t} a\left(\frac{\tau}{2}\right) G(s_0,\tau - s_0)\,d\tau + \\ & + \int_{2x_0}^{s_0+t} a\left(\frac{\tau}{2}\right)\cdot\big(G(s,\tau - s) - G(s_0,\tau - s_0)\big)\,d\tau \Bigg\} \\ =: & \int_{t_1}^{t_2} (X + Y + Z)\,dt. \end{aligned}$$

Für $\int_{t_1}^{t_2} X\,dt$ erhält man, wenn man den Mittelwertsatz der Differentialrechnung anwendet

$$\int_{t_1}^{t_2} X\,dt = \int_{t_1}^{t_2}\int_{s_0+t}^{s+t} a\left(\frac{\tau}{2}\right)\cdot(s - s_0)\big(G_s(s',\tau - s') - G_t(s',\tau - s')\big)\,d\tau\,dt,$$

wobei $s' = s'(\tau)$ stets zwischen s_0 und s liegt; da G_s und G_t um $(s_0)\times I_1$ beschränkt sind (z. B. auf C), $|G_s|, |G_t| \leqq K$, ist

$$\left|\frac{1}{s - s_0}\int_{t_1}^{t_2} X\,dt\right| \leqq 2K\left|\int_{t_1}^{t_2}\int_{s_0+t}^{s+t}\left|a\left(\frac{\tau}{2}\right)\right|d\tau\,dt\right|$$

und dieses Integral geht für $s \to s_0$ gegen 0, da das Maß des Parallelogrammes $\substack{t_1 \leqq t \leqq t_2 \\ s_0 + t \leqq \tau \leqq s + t}$ in der $\tau - t$-Ebene dann gegen 0 geht.

Eine analoge Anwendung des Mittelwertsatzes liefert unter Ausnutzung der gleichmäßigen Stetigkeit der G_s und G_t auf C

$$\lim_{s\to s_0}\frac{1}{s-s_0}\int_{t_1}^{t_2} Z\,dt \text{ existiert} = \int_{t_1}^{t_2}\int_{2s_0}^{s_0+t} a\left(\frac{\tau}{2}\right)\left(G_s(s_0,\tau-s_0)-G_t(s_0,\tau-s_0)\right)d\tau\,dt$$
$$=: z(s_0).$$

Für das mittlere Integral erhält man mit FUBINI

$$\int_{t_1}^{t_2} Y\,dt = \int_{t_1}^{t_2}\int_{s_0}^{s} a\left(\frac{s'+t}{2}\right)G(s_0,t+s'-s_0)\,ds'\,dt$$
$$= \int_{s_0}^{s}\int_{t_1}^{t_2} a\left(\frac{s'+t}{2}\right)G(s_0,t+s'-s_0)\,dt\,ds'$$
$$= \int_{s_0}^{s}\int_{t_1+s'}^{t_2+s'} a\left(\frac{\tau}{2}\right)G(s_0,\tau-s_0)\,d\tau\,ds' =: \int_{s_0}^{s} y(s')\,ds'.$$

Der Integrand $y(s')$ ist stetig bei $s'=s_0$, also existiert

$$\lim_{s\to s_0}\frac{1}{s-s_0}\int_{t_1}^{t_2} Y\,dt = y(s_0) = \int_{t_1+s_0}^{t_2+s_0} a\left(\frac{\tau}{2}\right)G(s_0,\tau-s_0)\,d\tau,$$

so daß schließlich auch $W(s)$ in s_0 differenzierbar ist und

$$W'(s)_0 = y(s_0) + z(s_0).$$

Da s_0 keinen Einschränkungen unterworfen war, ist $W(s)$ überall differenzierbar, die Darstellungen für y und z zeigen überdies, daß $W(s)$ sogar stetig differenzierbar ist. Nach (15) ist damit $U(s)$ und ebenso $V(t)$ überall zweimal stetig differenzierbar. Weiter ist, wenn man sich auf eine Umgebung des Punktes (s,t) aus B beschränkt und $s=s_0$ fest läßt, gemäß (6') $H_s(s_0,t)$ in t totalstetig und daher f.ü nach t dort differenzierbar[8]), so daß wegen $G_s = U' + \frac{1}{4}H_s$ dasselbe auch von $G_s(s_0,t)$ gilt; nach (12) wird nun

$$h'\left(\frac{s+t}{2}\right) = \exp\left(\frac{l}{2}(s-t)\right)\cdot\left(l\cdot G(s,t) + 2G_s(s,t)\right),$$

daher ist $h'\left(\frac{s+t}{2}\right)$ dort ebenfalls totalstetig und f.ü. nach t differenzierbar; weil nun (s_0,t) beliebig aus B war, folgt, wenn man $s_0=0$ und $t/2=x$ setzt, daß $h'(x)$ auf I lokal totalstetig und f.ü differenzierbar ist.

Damit läßt sich nun der Beweis von Satz 3.1 rasch zu Ende führen: Es wird nach (12)

$$G_s(s,t) = \frac{1}{2}\exp\left(l\frac{-s+t}{2}\right)\cdot\left(h'\left(\frac{s+t}{2}\right) - l\cdot h\left(\frac{s+t}{2}\right)\right),$$

[8]) Im Gegensatz zu $U(s)$ und $V(t)$ ist $H(s,t)$ und damit auch G und F im allgemeinen nur dann zweimal stetig differenzierbar, wenn $a(x)$ stetig ist.

für $s=0$ also

$$G_s(0,t)=\frac{1}{2}e^{lt/2}\left(h'\left(\frac{t}{2}\right)-l\cdot h\left(\frac{t}{2}\right)\right) \quad \text{für } \frac{t}{2}\in I,$$

somit f.ü.

$$G_{st}(0,t)=\frac{1}{4}e^{lt/2}\left(h''\left(\frac{t}{2}\right)-l^2 h\left(\frac{t}{2}\right)\right).$$

Andererseits ist nach (6') für $s=0$, t aus $2I$

$$G_{st}=\frac{1}{4}H_{st}=\frac{1}{4}a\left(\frac{t}{2}\right)G(0,t)=\frac{1}{4}a\left(\frac{t}{2}\right)e^{lt/2}h\left(\frac{t}{2}\right) \quad \text{f.ü. in } t,$$

so daß man für $t/2=x$, $x\in I$, bis auf eine Menge vom Maß 0 (die Stellen, an denen $e^{lt/2}$ verschwindet, sind höchstens abzählbar)

$$h''(x)=\left(l^2+a(x)\right)\cdot h(x), \quad \text{das ist (11)},$$

auf I hat. q.e.d.

Mit diesem Satz lassen sich nun sofort die invarianten eindimensionalen Moduln von $\mathfrak{F}$ bestimmen:

Satz 3.2. *Eine Funktion $F(x,y)$ ist genau dann elementar und aus $\mathfrak{E}$, wenn sie die Form*

$$F(x,y)=h_\omega(x)\cdot e^{i\omega y} \tag{16}$$

hat, wo ω beliebig reell ist, wo $h_\omega(x)$ im Sinne von Satz 3.1 Lösung von

$$h''(x)-\left(a(x)-\omega^2\right)\cdot h(x)=0 \tag{17}$$

ist und wo zusätzlich noch

$$\sup_{x\in I}|\varrho(x)\,h_\omega(x)|<\infty \tag{18}$$

ist. Ein derartiges F ist dann auch fastperiodisch aus $\mathfrak{F}$.

Beweis. Erfüllt F die obigen Bedingungen, so ist es elementare verallgemeinerte Lösung von (1) (Satz 3.1), außerdem ist dann $\|F\|_\varrho<\infty$; F ist daher aus $\mathfrak{E}$. Die Fastperiodizität solcher F folgt dann aus der der $e^{i\omega y}$. Ist umgekehrt F elementar und aus $\mathfrak{E}$, so hat es nach Satz 3.1 die Form (10); wegen $\|F\|_\varrho<\infty$ muß dabei zunächst l rein imaginär $=:i\omega$ sein, außerdem muß dann auch noch (18) gelten.

Elementare fastperiodische F aus $\mathfrak{E}$ sind also stetig, einmal stetig differenzierbar und (gleichgradig in x) y-periodisch.

Unter einer *Funktion* $h_\omega(x)$ werde hier und im folgenden stets eine Lösung von (17) verstanden, wobei Lösung wie in Satz 3.1 definiert ist. Für die *Lösungsmannigfaltigkeit* L_ω von (17) bei festem ω gelten dann, obwohl $a(x)$ nur als lokal integrabel (und komplexwertig) statt wie gewöhnlich stetig (und reell) vorausgesetzt wurde, die üblichen Sätze (vgl. [7], Anhang II), insbesondere gibt es zu jeder Anfangsbedingung genau eine Lösung, L_ω ist also zweidimensional.

Nach diesen Bemerkungen kann man den Approximationssatz für fastperiodische Vektoren aus $\mathfrak{E}$ formulieren:

Satz 3.3 (Approximationssatz). *Jeder abgeschlossene invariante lineare Teilraum $\mathfrak{F}_1$ von $\mathfrak{F}$ wird aufgespannt von den elementaren fastperiodischen $F(x, y)$ aus $\mathfrak{F}_1$, d.h. jedes $F(x, y)$ aus $\mathfrak{F}_1$ ist im Sinne der Norm von $\mathfrak{E}$ durch endliche Linearkombinationen von Funktionen $h_\omega(x) \cdot e^{i\omega y}$ aus $\mathfrak{F}_1$ (ω reell, h_ω wie in Satz 3.2) beliebig genau approximierbar.*

Ein $F(x, y)$ ist also genau dann fastperiodischer Vektor aus $\mathfrak{E}$, wenn es zu jedem positiven ε ein natürliches N, reelle Zahlen ω_m ($m = 1, \ldots, N$) und zugehörige Funktionen $h_{\omega_m}(x)$, die (18) erfüllen, gibt, so daß

$$\left\| F(x, y) - \sum_1^N h_{\omega_m}(x) \cdot e^{i\omega_m y} \right\|_\varrho < \varepsilon. \tag{19}$$

Der Beweis folgt sofort aus dem zu Beginn dieses Abschnittes erwähnten Approximationssatz in MAAK [12], § 2, in Verbindung mit unserem Satz 3.2.

Für gerades $F(x, y)$ läßt sich Satz 3.3 etwas verschärfen, dann sind auch die Approximationssummen gerade wählbar, also von der Form

$$\sum_m h_{\omega_m}(x) \cdot \cos \omega_m y.$$

Beispiele für Räume $\mathfrak{F}_1 \neq \mathfrak{F}$, auf die Satz 3.3 angewandt werden kann, sind die Mengen aller in y mit einer festen Periode τ periodischen Funktionen aus $\mathfrak{F}$, oder die Mengen aller derjenigen $F \in \mathfrak{F}$, die in $x = c$ und $x = d$ gewissen Randbedingungen genügen (d.s. solche Funktionen, auf die auch die Ergebnisse von BOCHNER und VON NEUMANN [3] angewandt werden können).

Zur Konvergenz in (19) für $\varepsilon \to 0$ ist zu bemerken, daß sie auf jedem abgeschlossenen Streifen endlicher Breite aus P gleichmäßig erfolgt, bei geeigneter Wahl von $\varrho(x)$ kann man auch gleichmäßige Konvergenz auf ganz P erreichen.

Überhaupt bestimmt die Normfunktion $\varrho(x)$ weitgehend den Umfang von $\mathfrak{F}$, sie kann z.B. so gewählt sein, daß die *Teilmannigfaltigkeit* l_ω von L_ω, die (18) genügt, nur eindimensional ist, oder daß $\mathfrak{F}$ nur solche Funktionen enthält, die auf den Geraden $x = x_i$ ($i \in J$) identisch verschwinden. Eine Abschwächung des Konvergenzbegriffes über den durch $\varrho_0(x)$ von Satz 1.4 bestimmten hinaus liefert aber keine neuen fastperiodischen Elemente mehr: Zwar folgt aus $\bigcup_\varrho \mathfrak{E}_\varrho = \mathfrak{E}_{\varrho_0}$ nicht, daß jedes $\mathfrak{F}_\varrho$ in $\mathfrak{F}_{\varrho_0}$ liegt, da aber jede einzelne ϱ-Norm gleichmäßig die ϱ_0-Norm beschränkt, sind ϱ-fastperiodische verallgemeinerte Lösungen von (1) auch ϱ_0-fastperiodisch; da weiter jede elementare Funktion $h_\omega(x)\, e^{i\omega y}$ mit $\varrho(x) := \left(1 + |h_\omega(x)|\right)^{-1}$ (eine zulässige Normfunktion!) fastperiodische verallgemeinerte Lösung von (1) aus $\mathfrak{F}_\varrho$ ist, hat man

Satz 3.4. *Die nach Satz 1.4 existierende universelle Normfunktion $\varrho_0(x)$ fängt auch alle fastperiodischen verallgemeinerten Lösungen von (1) ein,*

$$\bigcup_\varrho \mathfrak{F}_\varrho = \mathfrak{F}_{\varrho_0}.$$

Bei dieser Normfunktion existieren daher stets nicht-identisch verschwindende fastperiodische Vektoren, z. B. sind alle elementaren Funktionen $h_\omega(x) \cdot e^{i\omega y}$ in $\mathfrak{F}_{\varrho_0}$.

Nach diesem Satz gilt also für alle Lösungen $h_\omega(x)$ von (17) bei beliebigem reellem ω eine *Abschätzung* der Form

$$|h_\omega(x)| = 0\left((1+|x|)\cdot e^{|x-x_0|\cdot\int_{x_0}^{x}|a(x')|\,dx'}\right) \qquad x \to \text{Rand}\,(I). \tag{20}$$

§ 4. Einige Eigenschaften fastperiodischer Lösungen

Satz 4.1. *Jedes fastperiodische $F(x, y)$ aus $\mathfrak{F}$ ist auf P stetig; jedes solche F ist in y gleichmäßig stetig, und zwar „ϱ-gleichgradig auf I", d.h. zu jedem positiven ε gibt es ein positives δ, so daß für alle reellen y_1, y_2, für die $|y_1 - y_2| < \delta$, und alle x aus I gilt*

$$\varrho(x)\cdot|F(x, y_1) - F(x, y_2)| < \varepsilon\ ^{9}).$$

Ist $a(x)$ gleichmäßig lokal integrabel (s. u.), und hat $\varrho(x)$ auf I von 0 einen positiven Abstand, so ist ein fastperiodisches $F(x, y)$ sogar gleichmäßig stetig auf P.

Schon die punktweise Stetigkeit ist für allgemeine F aus $\mathfrak{E}$ im allgemeinen nicht erfüllt, erst recht nicht gelten die verschiedenen Arten der gleichmäßigen bzw. gleichgradigen Stetigkeit allgemein in $\mathfrak{E}$.

Da auf jedem kompakten Teilintervall I_1 von I $\int_{I_1}|a(x)|\,dx < \infty$ und $\varrho(x) \geqq \delta_{I_1} > 0$ ist, ist auf dem dazugehörigen Streifen $I_1 \times R$ ein fastperiodisches $F(x, y)$ stets gleichmäßig stetig.

Beweis. Nach Satz 3.1 sind die elementaren F stetig auf P, nach Satz 3.3 und weil die Approximation durch elementare F lokal gleichmäßig ist, sind dann auch beliebige fastperiodische F aus $\mathfrak{F}$ stetig auf P. Analog zeigt man die Richtigkeit der zweiten Aussage von Satz 4.1, da elementare fastperiodische F diese Eigenschaft haben.

Was die letzte Aussage von Satz 4.1 betrifft, so genügt es wegen Satz 3.3 und der über $\varrho(x)$ gemachten Voraussetzung, sie für elementare fastperiodische Vektoren nachzuweisen; solche F sind aber genau dann gleichmäßig auf P stetig, wenn das zugehörige $h(x)$ auf I gleichmäßig stetig ist. Die gleichmäßige Stetigkeit von $h(x)$ folgt jedoch, da wegen $\varrho \geqq \delta_0 > 0$ und $\sup_I \varrho \cdot |h| < \infty$ die Funktion $h(x)$ auf I beschränkt ist, aus dem folgenden

Hilfssatz. *Ist $h(x)$ eine auf I beschränkte Lösung*[10]) *von*

$$h''(x) = b(x)\cdot h(x),$$

wo $b(x)$ auf I meßbar und gleichmäßig lokal integrabel ist, so ist auch $h'(x)$ auf I beschränkt.

Dabei soll eine auf $I = \{c, d\}$ definierte meßbare Funktion $b(x)$ *auf I gleichmäßig lokal integrabel* heißen, falls es ein positives (festes) $\delta_0 < d - c$ gibt, so daß

$$\sup_{x,\, x+\delta_0 \in I}\int_x^{x+\delta_0}|b(t)|\,dt < \infty.$$

---

[9]) Anders ausgedrückt: Die Abbildung $T_t F: R \to \mathfrak{E}$ ist für fastperiodische F in t stetig.

[10]) Ist das Intervall I beschränkt und $b(x)$ über I integrabel, so ist jede Lösung $h(x)$ auf I beschränkt.

Beweis des Hilfssatzes. Ist I endlich, so ist auf Grund der über b gemachten Voraussetzung $\int_I |b|\,dt<\infty$, daher gilt

$$|h'(x)-h'(x_0)| = \left|\int_{x_0}^{x} h''(t)\,dt\right| = \left|\int_{x_0}^{x} b(t)\,h(t)\,dt\right| \leqq \sup_I |h| \cdot \int_I |b|\,dt<\infty.$$

Ist I unendlich (einseitig oder $=R$), so schließen wir indirekt: Ist $h'(x)$ nicht beschränkt, so gibt es zu jedem natürlichen n ein $x_n\in I$, so daß z.B. $h'(x_n)>n$; wegen

$$|h'(x)-h'(x_n)| = \left|\int_{x_n}^{x} b(t)\,h(t)\,dt\right| \leqq \sup_I |h| \cdot \left|\int_{x_n}^{x} |b|\,dt\right| =: A \cdot \left|\int_{x_n}^{x} |b|\,dt\right|$$

ist dann für $\left|\int_{x_n}^{x} |b|\,dt\right| \leqq \frac{n}{2A}$, $x\in I$ jedenfalls

$$h'(x)>\frac{n}{2}, \quad \text{dort also} \quad |h(x)-h(x_n)| \geqq \frac{n}{2}|x-x_n|;$$

durch passende Wahl von x (I ist unendlich!) kann man daher erreichen, daß für hinreichend große n

$$|h(x)-h(x_n)| \geqq \frac{n}{2}\delta_0,$$

d.h. $h(x)$ ist nicht beschränkt.

Satz 4.2. *Eine Funktion F aus $\mathfrak{E}$ ist genau dann fastperiodisch, wenn sie*

a) y-stetig ist, und zwar ϱ-gleichgradig in $x\in I$, d.h. zu jedem reellen y_1 und positiven ε gibt es ein positives δ, so daß, falls $|y-y_1|<\delta$, für alle $x\in I$

$$\varrho(x)\cdot|F(x,y)-F(x,y_1)|<\varepsilon$$

und

b) wenn es zu jedem positiven ε eine relativ dichte Menge $T\subset R$ von ε-Translationszahlen τ für F gibt, die ϱ-gleichgradig in $x\in I$ gelten, d.h. für beliebige $\tau\in T$ gilt

$$\|F-T_\tau F\|_\varrho<\varepsilon.$$

Eine Menge T von reellen Zahlen heißt dabei *relativ dicht*, wenn es eine reelle Zahl L gibt, so daß in jedem reellen Intervall der Länge L mindestens eine Zahl τ aus T liegt (vgl. [11], S. 92).

Es ist hier anzumerken, daß die Bedingung a) schwächer ist als die Forderung „T_tF ist eine in t stetige Abbildung von R in $\mathfrak{E}$".

Bei kompaktem Intervall I und beschränktem ϱ ist a) ersetzbar durch

a') $F(x,y)$ ist auf P stetig.

Beweis. Daß a) und b) notwendig sind, folgt aus Satz 4.1 und dem Approximationssatz 3.3 [denn elementare F sind sogar (gleichgradig in $x\in I$) periodisch in y, daraus erschließt man dann analog wie bei H. Bohr, Acta Math. **45**, 37–40 (1925), daß auch endliche Linearkombinationen solcher Funktionen Bedingung b) erfüllen].

Das Umgekehrte folgt aus einem allgemeinen Satz dieser Art: Seien G und E zwei topologische abelsche Gruppen[11]), $f(y)$ eine Funktion auf G mit Werten in E. Eine Teilmenge T von G heiße *relativ dicht* (in [6] wird sie *syndetic* genannt), wenn es eine kompakte Menge L aus G gibt, so daß für alle y aus G die Menge $y+L$ mit T einen nicht-leeren Durchschnitt hat (im Falle $G=R=$ additive Gruppe der reellen Zahlen stimmt dies mit der früheren Festsetzung überein; allgemein ist damit gleichwertig: Es gibt eine kompakte Menge K, so daß $K+T=G$).

Weiter werde τ *Translationszahl* für f zur Umgebung $U, O\in U\subset E$, genannt, wenn für alle y aus G

$$f(y)-f(y+\tau)\in U.$$

Es gilt dann der

Satz. *Ist $f(y)$ stetig, so ist f genau dann fastperiodisch (im Maakschen Sinne), wenn es zu jeder Umgebung der 0 in E eine relativ dichte Menge T von Translationszahlen für f gibt.*

Der Beweis findet sich in [6], S. 42 oder in [7], S. 36/37, er verläuft für die hier benötigte Richtung so, daß man zunächst die gleichmäßige Stetigkeit von $f(y)$ zeigt; teilt man dann, bei vorgegebener Umgebung $U\subset E$, die kompakte Menge $-L$ so in endlich viele Teilmengen $-L_1, \ldots, -L_N$, daß $f(y)$ auf ihnen hinreichend wenig variiert, so liefert $A_n =: -L_n+T$ $(n=1, \ldots, N)$ eine gesuchte Teilung von G für f.

Damit läßt sich nun leicht die zweite Behauptung von Satz 4.2 beweisen. Dazu setzen wir

$G := R$ (additive Gruppe der reellen Zahlen mit der üblichen Topologie),

$E :=$ Menge aller komplexwertigen Funktionen $\varphi(x)$, die auf I definiert sind und für die $|\varphi|_\varrho := \sup_{x\in I} \varrho(x)\cdot|\varphi(x)|<\infty$;

diese Menge wird durch $|\varphi|_\varrho$ und das übliche „+" zu einer topologischen abelschen Gruppe

$f(y) := F(x, y)$.

$f(y)$ ist dann eine Abbildung von G in E (da F aus $\mathfrak{E}$, ist $|f(y)|_\varrho$ stets endlich); man prüft sofort nach, daß f stetig ist [Bedingung a)], sowie daß es stets relativ dicht liegende Translationszahlen besitzt, so daß f also fastperiodisch ist. Die Fastperiodizität von $f(y)$ gibt aber genau die von $F(x, y)$ im früher (Abschnitt 2) definierten Sinne.

Im übrigen könnte man die Bedingung $\|F\|_\varrho<\infty$, die in $F\in\mathfrak{E}$ enthalten ist, für Satz 4.2 analog wie bei Satz 2.1 abschwächen, doch soll hierauf verzichtet werden.

§ 5. Fourier-Reihen

Hat man allgemein eine Funktion $f(t)$ auf einer Gruppe G mit Werten in einem Banach-Raum E, die fastperiodisch ist, so ist $f(t)$ ergodisch, d.h. es

[11]) Die Kommutativität ist hier nicht wesentlich, der folgende Satz gilt auch für nicht-abelsche Gruppen, wenn man die Begriffe *relativ dicht* und *Translationszahl* richtig formuliert.

besitzt einen Mittelwert mit den üblichen Eigenschaften (vgl. MAAK [12], § 1); allgemeiner gilt dies auch von allen Funktionen der Form $f(t) \cdot \bar{a}_{ij}(t)$, wo $f(t)$ fastperiodisch auf G und $(a_{ij}(t))$ eine endlich-dimensionale beschränkte Darstellung von G ist, da dann auch diese Ausdrücke fastperiodisch sind. Über diesen abstrakten Mittelwert $M_t(f(t) \cdot \bar{a}_{ij}(t)) =: A_{ij}$, der Element von E ist, kann man im vorliegenden Fall genauere Aussagen machen:

Satz 5.1. *Ist $F(x, y)$ fastperiodischer Vektor aus $\mathfrak{F}$, so existieren für jedes reelle ω die folgenden Mittelwerte und es gilt*

$$A_\omega(x, y) := M_t(T_t F \cdot e^{-i\omega t}) = \left[\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} F(x, y') e^{-i\omega y'} dy'\right] \cdot e^{i\omega y}.$$

Die dadurch definierten Vektoren $A_\omega(x, y)$ sind aus $\mathfrak{F}$ und elementar zu ω, also

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} F(x, y')\, e^{-i\omega y'} dy' = h_\omega(x)$$

gemäß Satz 3.2. Dabei existiert dieser Mittelwert ϱ-gleichgradig in x, d.h. zu jedem positiven ε gibt es ein T_ε, so daß für alle $T > T_\varepsilon$ und alle x aus I

$$\varrho(x) \cdot \left| h_\omega(x) - \frac{1}{2T} \int_{-T}^{T} F(x, y')\, e^{-i\omega y'} dy' \right| < \varepsilon.$$

Satz 5.1 besagt also unter anderem, daß abstrakter Mittelwert und Bohrscher Mittelwert hier übereinstimmen; wir schreiben daher später (noch nicht im folgenden Beweis)

$$M_y(f(y)) \quad \text{für} \quad \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} f(y)\, dy.$$

Beweis. Wegen der Invarianz des ergodischen Mittelwertes hat man

$$\begin{aligned} T_c A_\omega(x, y) &= A_\omega(x, y + c) = M_t(T_t F(x, y + c) \cdot e^{-i\omega t}) \\ &= M_t(F(x, y + c + t)\, e^{-i\omega t}) = M_{t'}(F(x, y + t') \cdot e^{-i\omega t'} \cdot e^{i\omega c}) \\ &= e^{i\omega c} \cdot A_\omega(x, y). \end{aligned}$$

Da außerdem $A_\omega(x, y)$ Element von $\mathfrak{F}$ ist, hat es nach Satz 3.2 die Form $h_\omega(x) \cdot e^{i\omega y}$.

Hält man in den Näherungssummen für den abstrakten Mittelwert x und y konstant, so erkennt man, daß $A_\omega(x, y)$ Mittelwert der Bohr-f.p. Funktion $f(c) = F(x, y + c) \cdot e^{-i\omega c}$ ist, wegen der eindeutigen Bestimmtheit eines solchen Mittelwertes (vgl. MAAK [11], S. 110) ist es der Bohrsche Mittelwert

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} f(c)\, dc.$$

Auf Grund der Invarianz dieses Mittelwerts wird daher

$$A_\omega(x,y) = \text{Bohr-}M_c\big(f(c)\big) = M_{y'}^B\left(F(x,y')\,e^{-i\omega y'}\right)\cdot e^{i\omega y}.$$

Ist $F \not\equiv 0$ selbst auch elementar, $F(x,y) = h_{\omega_0}(x)\,e^{i\omega_0 y}$, so gibt es sicher ein T_ε, so daß für $T > T_\varepsilon$ (ω beliebig reell, fest)

$$\left|\frac{1}{2T}\int_{-T}^{T} e^{i(\omega_0-\omega)y}\,dy - M_y^B\left(e^{i(\omega_0-\omega)y}\right)\right| < \varepsilon/\sup_{x\in I}\left|\varrho(x)\cdot h_{\omega_0}(x)\right|$$

oder

$$\varrho(x)\cdot\left|\frac{1}{2T}\int_{-T}^{T} F(x,y)\,e^{-i\omega y}\,dy - M_y^B\left(F(x,y)\,e^{-i\omega y}\right)\right| < \varepsilon \qquad \text{für alle } x \text{ aus } I.$$

Ein derartiges T gibt es dann auch, wenn F eine endliche Linearkombination elementarer f.p. Funktionen aus $\mathfrak{F}$ ist; da sich jedes f.p. F aus $\mathfrak{F}$ durch solche Linearkombinationen im Norm-Sinne beliebig genau approximieren läßt überträgt sich die Richtigkeit der obigen Ungleichung auf ganz $\mathfrak{F}$. q.e.d.

Definition 4. Die Funktionen $A_\omega(x,y)$ von Satz 5.1 heißen die *Fourier-Komponenten von* $F(x,y)$ *zu* ω (reell),

$$A_\omega(x,y) := M_y\left(F(x,y)\,e^{-i\omega y}\right)\cdot e^{i\omega y} = h_\omega(x)\cdot e^{i\omega y}. \tag{21}$$

Die formale Summe $\sum_\omega A_\omega(x,y)$ heißt die (abstrakte) *Fourier-Reihe von* $F(x,y)$,

$$F(x,y) \sim \sum_{\omega\ \text{reell}} h_\omega(x)\cdot e^{i\omega y}. \tag{22}$$

Fourier-Reihen heißen *gleich*, wenn sie gliedweise übereinstimmen.

Nun ist für ein Element $F\in\mathfrak{F}$ bei konstantem x die Funktion $F(x,y)$ Bohr-fastperiodisch und besitzt eine Bohrsche Fourier-Reihe $\sum_\omega a_\omega(x)\,e^{i\omega y}$, wo $a_\omega(x) = M_y\left(F(x,y)e^{-i\omega y}\right)$, nach Satz 5.1 schließen sich also die Koeffizienten $a_\omega = a_\omega(x)$ zu Lösungen $h_\omega(x)$ der Differentialgleichung (17) zusammen,

$$a_\omega(x) = h_\omega(x);$$

kurz ausgedrückt:

Satz 5.2. *Die abstrakte Fourier-Reihe von F ist identisch mit der Bohrschen Fourier-Reihe von F.*

Die Parseval-Gleichung für Bohr-fastperiodische Funktionen (MAAK [11], S. 117) nimmt hier die Form an

$$M_y\left(|F(x,y)|^2\right) = \sum_{\omega\ \text{reell}} |h_\omega(x)|^2 = \sum_{\omega\ \text{reell}} |A_\omega(x,y)|^2$$

[die wegen $M_y\left(|F(x,y)|^2\right) = M_t\left(|T_t F(x,y)|^2\right)$ auch abstrakt gedeutet werden kann]:

Satz 5.3. *Für fastperiodische Vektoren F aus $\mathfrak{F}$ gilt eine Parsevalsche Gleichung*

$$M_y\left(|F(x,y)|^2\right) = \sum_{\omega\ \text{reell}} |A_\omega(x,y)|^2 \qquad x \text{ aus } I. \tag{23}$$

Daraus kann man die üblichen Folgerungen über Fourier-Reihen ziehen:

Zunächst erhält man durch Integration über ein beliebiges, festes, kompaktes Teilintervall I_1 von I [dort ist $F(x, y)$ beschränkt]

$$\sum_1^m \int_{I_1} |h_{\omega_\nu}(x)|^2 dx \leqq K_{I_1} < \infty$$

für jedes endliche System reeller $\omega_1, \ldots, \omega_m$, so daß nur höchstens abzählbar viele $\int_{I_1} |h_\omega(x)|^2 dx$ von Null verschieden sein können und damit auch nur höchstens abzählbar viele $A_\omega(x, y)$ auf $I \times R$[12]).

Haben F_1 und F_2 dieselbe Fourier-Reihe, so hat $F_1 - F_2$ die Fourier-Reihe $\equiv 0$ [nach (21) darf man Fourier-Reihen linear kombinieren], also ist nach (23) $M_y(|F_1 - F_2|^2) = 0$ auf I. Da der Mittelwert einer nicht-negativen Bohr-fastperiodischen Funktion nur dann verschwindet, wenn die Funktion selbst verschwindet (MAAK [11], S. 42), hat man $F_1 - F_2 = 0$, d. h. die Fourier-Reihe bestimmt $F(x, y)$ eindeutig.

Satz 5.4. *In einer Fourier-Reihe* (22) *sind nur höchstens abzählbar viele $A_\omega(x, y)$ bzw. $h_\omega(x)$ nicht identisch Null. Fourier-Reihen darf man linear kombinieren. Fastperiodische $F(x, y)$ aus $\mathfrak{F}$ mit gleicher Fourier-Reihe sind identisch auf P, die Beziehung zwischen den Vektoren F und ihren Fourier-Reihen ist also eineindeutig.*

Es gibt nun noch einen etwas anderen Weg, zu einer Fourier-Reihe einer fastperiodischen Funktion, die ihre Werte in einem Banach-Raum hat, zu kommen:

Statt F mit den Darstellungskoeffizienten der vorliegenden Gruppe zu mitteln, was Fourier-Komponenten ergibt, die Vektoren in $\mathfrak{F}$ sind, kann man auch ein inneres Produkt (F_1, F_2) zwischen fastperiodischen Funktionen erklären und kommt dann, wenn man speziell (F, F_{el}) bildet, wo F_{el} elementare fastperiodische Funktion auf der gegebenen Gruppe ist, die ihre Werte in demselben Banach-Raum hat wie F, zu Fourier-Koeffizienten, die komplexe Zahlen sind (vgl. MAAK [12], § 3).

Dazu muß zunächst ein inneres Produkt definiert werden: Sei $\sigma(x)$ eine auf I definierte *Gewichtsfunktion*, deren Werte positive reelle Zahlen sind, die meßbar ist und lokal von 0 einen positiven Abstand hat und für die

$$\int_I \frac{\sigma(x)}{(\varrho(x))^2} dx < \infty$$[13]).

Derartige $\sigma(x)$ gibt es für jedes zulässige $\varrho(x)$ (§ 1), und zwar in mannigfacher Art; durch Normierung kann man noch erreichen, daß obiges Integral zu 1 wird. Irgendein solches $\sigma(x)$ werde für das folgende fest ausgewählt. Statt $\int_I \sigma(x) C(x) dx$ schreiben wir dann im folgenden kurz $I_x(C(x))$.

[12]) Ist $h_\omega(x) = 0$ f. ü. auf I_1, so verschwindet es auf ganz I.

[13]) Sollte $\varrho(x)$ nicht meßbar sein, so kann man es hier und in den folgenden Abschätzungen stets so durch ein meßbares $\tilde{\varrho}(x)$ ersetzen, daß die späteren Sätze richtig bleiben [und zwar für das ursprünglich gegebene $\varrho(x)$].

Definition 5. Für fastperiodische Vektoren F, G aus $\mathfrak{F}$ sei

$$(F, G)_\sigma := I_x\left(M_y\left(F(x, y) \cdot \overline{G(x, y)}\right)\right) \tag{24}$$

sowie

$$|||F|||_\sigma := \sqrt{(F, F)_\sigma}.$$

Zunächst hat man hier

Satz 5.5. *Für fastperiodische F und G aus $\mathfrak{F}$ existiert $M_y\left(F(x, y)\,\overline{G(x, y)}\right)$ ϱ^2-gleichgradig auf I (vgl. Satz 5.1); es gilt $\sigma \cdot M_y(F\overline{G}) \in \mathfrak{L}^1(I)$* sowie

$$|(F, G)_\sigma| \leqq \|F\|_\varrho \cdot \|G\|_\varrho,$$

das Skalarprodukt ist also stets endlich.

Die beiden Mittelwerte sind darin vertauschbar,

$$(F, G)_\sigma = I_x M_y(F\overline{G}) = M_y I_x(F\overline{G}).$$

$I_x(F\overline{G})$ ist Bohr-fastperiodisch.

Der Beweis, der teilweise analog wie der von Satz 5.1 verläuft, kann hier ebenso wie der des folgenden Satzes übergangen werden (vgl. [7], S. 43/44).

Satz 5.6. *Der Raum $\mathfrak{F}$ der fastperiodischen Vektoren wird, versehen mit dem inneren Produkt* (24), *zu einem euklidischen Raum (d. i. ein Hilbert-Raum bis auf die Vollständigkeit in der $|||_\sigma$-Topologie). Die Gruppe $\mathfrak{G}$ der Translationen T_c ist unitär, also $(T_c F, G)_\sigma = (F, T_{-c} G)_\sigma$, so daß die Norm $|||F|||_\sigma$ invariant gegenüber $\mathfrak{G}$ ist. Außerdem gilt*

$$|||F|||_\sigma \leqq \|F\|_\varrho.$$

Zu bemerken ist hier, daß die Einführung der $|||_\sigma$-Norm in $\mathfrak{F}$ die bisherige Gleichheitsdefinition zwischen Elementen von $\mathfrak{F}$ (nämlich Identität) nicht ändert, aus $|||F|||_\sigma = 0$ folgt also $F \equiv 0$ auf P.

Nach Satz 5.6 sind jetzt alle Voraussetzungen aus MAAK [12], § 3 erfüllt, so daß man gemäß den dortigen Überlegungen auch auf die folgende Art zu einer Fourier-Reihe für fastperiodische Elemente von $\mathfrak{F}$ gelangen kann:

Wir betrachten die endlich-dimensionalen irreduziblen invarianten Teilmoduln von $\mathfrak{F}$, die zu einem festen ω gehören (nur solche können zu äquivalenten Darstellungen führen); diese Moduln spannen einen *invarianten Teilmodul* $\mathfrak{F}_\omega$ von $\mathfrak{F}$ auf, der nach Satz 3.2 und den darauf folgenden Bemerkungen höchstens zweidimensional ist. In jedem dieser Teilmoduln wählen wir eine Basis aus, die bezüglich des inneren Produktes (24) orthonormiert ist und bezeichnen die diese Basis bildenden Funktionen mit

$$F_{\omega, v}(x, y) = h_{\omega, v}(x) \cdot e^{i\omega y} \qquad (v = 1, 2)$$

(ist $\mathfrak{F}_\omega$ 0- oder 1-dimensional, so setze man $h_{\omega, v}$ bzw. $h_{\omega, 2} = 0$). Die $F_{\omega, v} \neq 0$ bilden dann eine orthonormierte Basis in $\mathfrak{F}$.

Ist F fastperiodischer Vektor, so heißen gemäß MAAK [12] die Zahlen

$$a_{\omega, v} := (F, F_{\omega, v})_\sigma \qquad (v = 1, 2;\ \omega \text{ reell})$$

σ-Fourier-Koeffizienten von F, während

$$F \sim \sum_{\omega \text{ reell}} (a_{\omega,1} F_{\omega,1} + a_{\omega,2} F_{\omega,2}) \tag{25}$$

σ-Fourier-Reihe von F heißt. Diese Ausdrücke hängen zunächst scheinbar von $(.,.)_\sigma$ und damit von $\sigma(x)$ ab (außerdem noch von der gewählten Basis).

Satz 5.7. *Die σ-Fourier-Reihe* (25) *stimmt überein mit der früher definierten Fourier-Reihe* (22), *genauer gilt*

$$A_\omega(x, y) = a_{\omega,1} F_{\omega,1}(x, y) + a_{\omega,2} F_{\omega,2}(x, y)$$

auf P für alle reellen ω.

In diesem Sinne sind die σ-Fourier-Koeffizienten unabhängig von σ und von der gewählten Basis[14]).

Zum Beweis stellen wir zuerst fest, daß auf Grund der Orthonormierung der $F_{\omega,\nu}$ i. a. $I_x(h_{\omega,i} \cdot \overline{h_{\omega,j}}) = \delta_{ij}$ gilt.

Nun läßt sich jedes einer Fourier-Komponente (21) zugeordnete $h_\omega(x)$ darstellen in der Form (ist $h_{\omega,\nu} = 0$, so kann man auch $A_{\omega,\nu} = 0$ annehmen)

$$h_\omega(x) = A_{\omega,1} h_{\omega,1}(x) + A_{\omega,2} h_{\omega,2}(x).$$

Multipliziert man diese Gleichung mit $\overline{h_{\omega,\nu}(x)}$ und mittelt über x, so erhält man

$$I_x\left(h_\omega(x)\,\overline{h_{\omega,\nu}(x)}\right) = A_{\omega,\nu}$$

oder mit (21)

$$A_{\omega,\nu} = I_x\left\{M_y\left(F(x, y)\, e^{-i\omega y}\right) \cdot \overline{h_{\omega,\nu}(x)}\right\} = (F, F_{\omega,\nu})_\sigma = a_{\omega,\nu},$$

womit die Übereinstimmung gezeigt ist.

Es sind also auch nur höchstens abzählbar viele $a_{\omega,\nu} \neq 0$, diese σ-Fourier-Koeffizienten bestimmen, bei gegebener fester Basis, $F(x, y)$ eindeutig.

Aus dem in Maak [12], § 3 Bewiesenen folgt hier noch

Satz 5.8. *Die Fourier-Reihe von $F \in \mathfrak{F}$ konvergiert im $|||_\sigma$-Norm-Sinne gegen F und es gilt eine Parseval-Gleichung*

$$|||F|||_\sigma^2 = \sum_{\omega \text{ reell}} \left(|a_{\omega,1}|^2 + |a_{\omega,2}|^2\right). \tag{26}$$

[Diese Form der Parsevalschen Gleichung erhält man übrigens auch sofort durch Integration von (23).]

Zum Abschluß noch eine Charakterisierung gerader fastperiodischer Funktionen $F(x, y)$: Ein Element $F \in \mathfrak{F}$ ist genau dann gerade, wenn für die zugehörigen Fourier-Komponenten gilt

$$h_\omega(x) = h_{-\omega}(x).$$

---

[14]) Die Unabhängigkeit der in Maak [12], § 3 eingeführten Fourier-Reihe von dem speziell verwendeten inneren Produkt und von der Basis gilt auch noch für allgemeinere Gruppen $\mathfrak{G}$ und Räume $\mathfrak{E}$.

§ 6. Beispiele

In diesem Abschnitt werden Beispiele zu den vorhergehenden Paragraphen untersucht; insbesondere werden die elementaren Funktionen für spezielle $a(x)$, also i.w. die Lösungen von

$$h''(x) - (a(x) - \omega^2) \cdot h(x) = 0 \tag{17}$$

angegeben. Bei periodischem $a(x)$ handelt es sich dabei um Differentialgleichungen vom Hillschen Typ, unter die eine Reihe wichtiger Differentialgleichungen fallen (z.B. Legendresche Differentialgleichung, konfluente hypergeometrische Differentialgleichung – vgl. KAMKE [8]).

1) $a(x) = 0$, $I = (-\infty, \infty)$, $\varrho(x) \equiv 1$.

Eine verallgemeinerte Lösung von (1) hat hier gemäß (3) die Form

$$F(x, y) = U(x - y) + V(x + y) \tag{27}$$

mit lokal integrablen U und V (die i.w. durch F eindeutig bestimmt sind). Ein solches F ist genau dann aus $\mathfrak{F}$, wenn $U(s)$ und $V(t)$ beschränkt sind.

Nach Satz 3.2 haben die elementaren fastperiodischen Funktionen die Gestalt

$$F_\omega(x, y) = a \cdot e^{i\omega(-x+y)} + b \cdot e^{i\omega(x+y)} \quad (\omega \text{ reell, } a, b \text{ komplex, beliebig}). \tag{28}$$

Mit Hilfe des Approximationssatzes 3.3 zeigt man:

Ein $F(x, y)$ ist genau dann fastperiodischer Vektor von $\mathfrak{F}$, wenn in (27) $U(s)$ und $V(t)$ Bohr-fastperiodisch sind.

(Im Falle eines endlichen Intervalls I gilt dies übrigens nicht mehr.)

2) $a(x) = A = \text{const}$.

[Die Differentialgleichung (1) ist dann eine Telegraphengleichung, A ist i.w. $= (RC - LG)^2$.]

Die elementaren fastperiodischen Lösungen haben hier gemäß Satz 3.2 das Aussehen

$$F_\omega(x, y) = a \cdot e^{\sqrt{A - \omega^2} \cdot x + i\omega y} + b \cdot e^{-\sqrt{A - \omega^2} \cdot x + i\omega y} \tag{29}$$
$$(\omega \text{ reell, } a, b \text{ komplex}),$$

je nach Art von I und $\varrho(x)$ kommen noch zusätzliche Bedingungen über ω und A hinzu, z.B. für $I = (-\infty, \infty)$, $\varrho(x) \equiv 1$ muß A reell sein und es kommen nur solche ω in Frage, für die $\omega^2 \geqq A$. Da die Phasengeschwindigkeit der elementaren Wellen gemäß (29) hier frequenzabhängig ist (x ist bis auf eine konstante Dehnung die Zeit t), folgt mit dem Approximationssatz, daß eine fastperiodische, also insbesondere periodische Spannungsverteilung längs der Leitung im Laufe der Zeit im allgemeinen ihre Form ändert, lediglich im „verzerrungsfreien Fall" $A = 0$ oder $R/G = L/C$ ist eine Übertragung ohne Verzerrungen möglich.

3) $a(x) = -x, \quad I = [0, +\infty)$.

Nach KAMKE [8], S. 440, (10), sind elementare verallgemeinerte Lösungen von (1) von der Form

$$F_\omega(x, y) = k(x + \omega^2) \cdot e^{i\omega y}, \tag{30}$$

wobei

$$k(t) := \sqrt{t} \cdot Z_{\frac{1}{3}}(\tfrac{2}{3} t^{\frac{3}{2}}), \qquad Z_{\frac{1}{3}} \text{ Zylinderfunktion.}$$

Durch entsprechende Wahl der Normfunktion $\varrho(x)$ bei 0 kann man die Dimension der von den $h_\omega(x) = k(x + \omega^2)$ bzw. F_ω gebildeten linearen Räume l_ω bzw. $\mathfrak{F}_\omega$ zu 1 oder 2 machen (unabhängig von ω).

4) $a(x) = \frac{k}{x^2}, \quad I = (0, +\infty)$.

Elementare fastperiodische $F(x, y)$ werden hier, wie man leicht nachprüft, durch folgende Ausdrücke gegeben

$$F_\omega(x, y) = z_k(|\omega|\, x) \cdot e^{i\omega y} \tag{31}$$

mit

$$z_k(t) := \sqrt{t} \cdot Z_{\sqrt{k+\frac{1}{4}}}(t), \qquad \operatorname{Re}\left(\sqrt{k + \tfrac{1}{4}}\right) \geqq 0, \qquad \omega \text{ reell}^{15});$$

$Z_\nu(t)$ ist dabei eine Zylinderfunktion, die der Besselschen Differentialgleichung genügt (vgl. MAGNUS-OBERHETTINGER [13], S. 25) und die in (31) so zu wählen ist, daß man für F eine endliche Norm erhält:

Bei gegebenem komplexem k kann man durch passende Wahl der Normfunktion $\varrho(x)$ (bei 0) erreichen, daß entweder nur die Besselschen Funktionen $J_{\sqrt{k+\frac{1}{4}}}$ oder aber auch noch die Neumannschen Funktionen $N_{\sqrt{k+\frac{1}{4}}}$ endliche Norm liefern [ersteres ist für $k \notin (-\infty, -\frac{1}{4})$ und $\varrho(x) = 1/\sqrt{x}$ der Fall]; auch hier ist die Dimension der einem ω zugeordneten Lösungsmannigfaltigkeit l_ω unabhängig von ω.

(Die von DELSARTE in [5] untersuchte Differentialgleichung

$$\Phi_{rr} + \frac{2p+1}{r} \Phi_r = \Phi_{tt}, \quad p \text{ fest reell}, \quad |p| < \frac{1}{2} \tag{32}$$

läßt sich durch die Transformation

$$r = x, \quad t = y, \quad F(x, y) = x^{\frac{2p+1}{2}} \cdot \Phi(x, y)$$

auf Beispiel 4 mit

$$k = \frac{4p^2 - 1}{4} \quad \left(-\frac{1}{4} \leqq k < 0\right) \quad \text{und} \quad \varrho(x) = x^{-\frac{2p+1}{2}}$$

zurückführen — fastperiodische Lösungen gehen bei dieser Transformation in ebensolche über.)

[15]) Der Faktor $\sqrt{t}$ ist hier durchaus natürlich: Die Zylinderfunktionen $Z_\nu(t)$ verschwinden im Unendlichen alle wie e^{-t}.

Für $\omega = 0$ sind übrigens die beiden $\mathfrak{F}_0$ bzw. L_0 in diesem Falle aufspannenden elementaren Funktionen gegeben durch

$$F_0(x, y) = \exp\left[\left(\tfrac{1}{2} \pm \sqrt{k + \tfrac{1}{4}}\right) \cdot \ln x\right] \tag{31'}$$

bzw. für $k = -\frac{1}{4}$: $F_0(x, y) = \sqrt{x}$ und $\sqrt{x} \cdot \ln x$.

5) $a(x) = -e^{2x}$, $I = (-\infty, +\infty)$.

Lösungen von (17) sind hier die Funktionen $h_\omega(x) = Z_{i\omega}(e^x)$, die zugehörigen elementaren $F_\omega(x, y)$ sehen so aus:

$$F_\omega(x, y) = Z_{i\omega}(e^x) \cdot e^{i\omega y}. \tag{33}$$

Literatur

[1] BEREZANSKIJ, M.: Über verallgemeinerte fastperiodische Funktionen und Formen, die mit Differential- und Differenzengleichungen zusammenhängen. Mat. Sbornik, n. Ser. **32** (74), 157–194 (1953) [Russisch]. — [2] BOCHNER, S.: Fastperiodische Lösungen der Wellengleichung. Acta math. **62**, 227–237 (1934). — [3] BOCHNER, S., and J. v. NEUMANN: On compact solutions of operational-differential equations. Ann. of Math., Ser. II **36**, 255–291 (1935). — [4] BOURBAKI, N.: Topologie Générale, Chap. X. Paris 1949. — [5] DELSARTE, J.: Une extension nouvelle de la théorie des fonctions presque-périodiques de BOHR. Acta math. **69**, 259–317 (1938). — [6] GOTTSCHALK, W., and C. HEDLUND: Topological Dynamics. Amer. Math. Soc. Coll. Publ. XXXVI (1955). — [7] GÜNZLER, H.: Hyperbolische Differentialgleichungen und Klassen fastperiodischer Funktionen. Diss. Univ. München 1957. — [8] KAMKE, E.: Differentialgleichungen, Lösungsmethoden und Lösungen. I, gewöhnliche Differentialgleichungen. Leipzig 1942. — [9] KAMKE, E.: Das Lebesgue-Stieltjes Integral. Leipzig 1956. — [10] LEVITAN, M. B.: Verallgemeinerte fastperiodische Funktionen. Mat. Sbornik, n. Ser. **24** (66), 321–346 (1949) [Russisch]. — [11] MAAK, W.: Fastperiodische Funktionen. Berlin 1950. — [12] MAAK, W.: Fastperiodische invariante Vektormoduln in einem metrischen Vektorraum. Math. Ann. **122**, 157–166 (1950). — [13] MAGNUS, W., u. F. OBERHETTINGER: Formeln und Sätze für die speziellen Funktionen der Mathematischen Physik. Berlin 1948. — [14] MARCENKO, V. A.: Verallgemeinerte fastperiodische Funktionen. Doklady Akad. Nauk SSSR., n. Ser. **74**, 893–895 (1950) [Russisch]. — [15] MUCKENHOUPT, C. F.: Almost-periodic functions and vibrating systems. J. Math. Physics **8**, 163–199 (1928/29).

Göttingen, Mathematisches Institut der Universität

(Eingegangen am 1. November 1958)

HARTMAN, P.
Math. Zeitschr. 71, 251–257 (1959)

On solutions of $\Delta V + V = 0$ in an exterior region

By
PHILIP HARTMAN[1])

HERGLOTZ has shown that if $V(x)$ is an entire solution of $\Delta V + V = 0$ subject to a boundedness condition, as $|x| \to \infty$, then certain "local means" of V satisfy an asymptotic relation, as $|x| \to \infty$; cf. [3]. It will be shown in Sections 1 to 2 that V itself satisfies the corresponding asymptotic relation in an averaged L^2-sense. This result is established quite easily and, in many applications, is more convenient than the result of HERGLOTZ. As an application, a very simple proof of a theorem of MAGNUS [2] is given in Section 3; cf. a proof by MÜLLER [3] using the Herglotz result. In Sections 4 to 7, it is shown that an analogue of the asymptotic formula obtained for entire solutions is valid for solutions $V = V(x)$ of $\Delta V + V = 0$ defined only for large $|x|$.

1. *An asymptotic relation.* Let $x = (x^1, \ldots, x^p)$, $p \geqq 2$, Δ be the (Euclidean) Laplacian and $V = V(x)$ a solution (of class C^2) of

$$\Delta V + V = 0 \tag{1}$$

in the entire x-space. The function V has for fixed r, a Fourier expansion in terms of spherical functions of the form

$$r^{\frac{1}{2}(p-1)} V(x) \sim (2\pi r)^{\frac{1}{2}} \sum_{n=0}^{\infty} i^n c_n J_\mu(r) K_n(u), \tag{2}$$

where $r = |x|$, $u = x/|x|$, c_n is a constant, J_μ is the Bessel function of order μ,

$$\mu \equiv \mu(n, p) = n + \tfrac{1}{2}(p-2), \tag{3}$$

$K_n(u)$ is a normalized spherical function of order n,

$$\int_U |K_n(u)|^2 \, du = 1, \tag{4}$$

$U: |u| = 1$ is the surface of the unit sphere and du is the element of area on U. (Below $dx = r^{p-1} \, du \, dr$ denotes the element of volume in x-space.)

If the asymptotic relation

$$(2\pi r)^{\frac{1}{2}} J_\mu(r) = 2\cos(r - \tfrac{1}{2}\mu\pi - \pi/4) + o(1), \qquad r \to \infty, \tag{5}$$

is substituted into (2) and use is made of (3) and of the fact that $K_n(-u) = (-1)^n K_n(u)$, there results formally an approximation for $r^{\frac{1}{2}(p-1)} V(ru)$ for

[1]) This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under contract No. AF 18 (603)-41. Reproduction in whole or in part is permitted for any purpose of the United States Government.

large r given by

$$W(ru) = i^{\frac{1}{2}(1-p)} e^{ir} K(u) + i^{\frac{1}{2}(p-1)} e^{-ir} K(-u), \tag{6}$$

where

$$K(u) \sim \sum_{n=0}^{\infty} c_n K_n(u). \tag{7}$$

HERGLOTZ has pointed out that, in general, one does not have $r^{\frac{1}{2}(p-1)} V(ru) - W(ru) = o(1)$ as $r \to \infty$ even under a boundedness condition

$$r^{-1} \int_{|x| \leqq r} |V(x)|^2 dx \leqq C \qquad \text{for all } r, \tag{8}$$

but that (8) implies that certain "local means" V_ε and W_ε, of V and W *do* satisfy $r^{\frac{1}{2}(p-1)} V_\varepsilon(ru) - W_\varepsilon(ru) = o(1)$ as $r \to \infty$; cf. [3].

The following theorem will be proved in the next section:

(I) *If $V = V(x)$ is a solution of* (1) *for all x satisfying* (8) *(for example, if $r^{\frac{1}{2}(p-1)} V(ru)$ is bounded), then*

$$r^{-1} \int_0^r \| t^{\frac{1}{2}(p-1)} V(tu) - W(tu) \|^2 dt \to 0 \quad \text{as} \quad r \to \infty, \tag{9}$$

where W is defined by (2), (6) *and* (7) *and $\| \ldots \|$ denotes the L^2-norm on U,*

$$\| V(ru) \|^2 = \int_U |V(ru)|^2 du, \tag{10}$$

and (7) *is understood in the L^2-sense.*

2. *Proof of* (I). PARSEVAL'S relation, in view of (2), gives for a fixed r,

$$r^{p-1} \| V(ru) \|^2 = 2\pi r \sum_{n=0}^{\infty} |c_n J_\mu(r)|^2. \tag{11}$$

If this relation is integrated over the interval $[0, r]$, it follows from (5) that every partial sum of $2 \sum |c_n|^2$ is majorized by C. Consequently,

$$\sum_{n=0}^{\infty} |c_n|^2 < \infty. \tag{12}$$

[Conversely, as pointed out by MAGNUS, (8) can be deduced from (12) by using

$$r^{-1} \int_0^r t J_\mu^2(t) dt \leqq \text{const.} \quad \text{for all } r, \ \mu \geqq 0. \tag{13}$$

The uniform boundedness of the expression on the left of (13) for $0 \leqq r < \infty$ and for bounded $\mu (\geqq 0)$ is clear from the corresponding uniformity of the $o(1)$ in (5). The fact that the inequality in (13) is valid for all $r, \mu \geqq 0$ follows from the recurrence relations for J_μ; cf. [2], pp. 85–86.]

When (12) holds, the theorem of FISCHER-RIESZ implies that (7) defines a function of $K(u)$ of class L^2 on the surface U of the unit sphere. The relation (6) gives a function $W(x) = W(ru)$ of class L^2 on any compact set in x-space.

By PARSEVAL'S relation,

$$\| r^{\frac{1}{2}(p-1)} V(ru) - W(ru) \|^2 = \sum_{n=0}^{\infty} |c_n|^2 |(2\pi r)^s J_\mu(r) - 2 \cos|^2,$$

where the argument of cos is the same as in (5). Clearly, term-by-term integration is permissible, so that the expression on the left of (9) is

$$\sum_{n=0}^{\infty} |c_n|^2 r^{-1} \int_0^r |(2\pi t)^{\frac{1}{2}} J_\mu(t) - 2\cos|^2 dt. \tag{14}$$

Since $|(2\pi t)^{\frac{1}{2}} J_\mu(t) - 2\cos|^2 < 4\pi t J_\mu^2(t) + 4$, it follows from (13) that the factor of $|c_n|^2$ in (14) is uniformly bounded for $0 \leqq r < \infty$ and $n = 0, 1, \ldots$. Hence (12) implies that the series is uniformly convergent on $0 \leqq r < \infty$. Since, by (5), each term tends to 0, as $r \to \infty$, the relation (9) follows.

3. *A theorem of Magnus.* The following theorem is given by MAGNUS in [2]:

(II) *If $V = V(x)$ is a solution of* (1) *for all x satisfying* (8) *and if the relation*

$$r^{\frac{1}{2}(p-1)} V(ru) \to 0 \quad \text{as} \quad r \to \infty \tag{15}$$

holds for every (fixed) u belonging to a half-sphere U^+, then $V(x) \equiv 0$.

By a half-sphere U^+ is meant a measurable subset of points u of the unit sphere U with the property that if U^- is the set of points $-u$, where u is in U^+, then, up to zero sets, U^+ and U^- are disjoint and their union is U.

In order to deduce (II) from (I), note that (6) shows that there exists an interval $[\alpha, \beta]$ in $[0, 2\pi]$ such that if $r \bmod 2\pi$ is in $[\alpha, \beta]$, then

$$\int_{U^+} |W(ru)|^2 du = \int_{U^+} (|K(u)|^2 + |K(-u)|^2)\, du + 2\,Re\, e^{2ir} i^{1-p} \int_{U^+} K(u)\, \overline{K}(-u)\, du$$

satisfies

$$\int_{U^+} |W(ru)|^2 du \geqq \int_{U^+} (|K(u)|^2 + |K(-u)|^2)\, du = \|K\|^2. \tag{16}$$

In view of (9), there exists an unbounded sequence $0 < r_1 < r_2 < \cdots$ with the property that $r_m \bmod 2\pi$ is in $[\alpha, \beta]$ and

$$\|r^{\frac{1}{2}(p-1)} V(ru) - W(ru)\| \to 0 \quad \text{as} \quad r = r_m \to \infty. \tag{17}$$

Since $|W(ru)|^2 \leq 2|K(u)|^2 + 2|K(-u)|^2$, there exists for every $\varepsilon > 0$, a $\delta = \delta_\varepsilon > 0$ (independent of r) such that if S is a measurable subset of U, then

$$\int_S |W(ru)|^2 du < \varepsilon \quad \text{whenever} \quad \int_S du < \delta_\varepsilon.$$

The last two relations show that

$$\int_S |r^{\frac{1}{2}(p-1)} V(ru)|^2 du < 2\varepsilon \quad \text{whenever} \quad \int_S du < \delta_\varepsilon$$

and $r = r_m$ is sufficiently large. Hence, EGOROFF's theorem and (15) imply that

$$\int_{U^+} |r^{\frac{1}{2}(p-1)} V(ru)|^2 du \to 0 \quad \text{as} \quad r = r_m \to \infty.$$

Thus, by (17),

$$\int_{U^+} |W(ru)|^2 du \to 0 \quad \text{as} \quad r = r_m \to \infty.$$

Since (16) holds for $r = r_m$, it follows that $\|K\| = 0$, that is, that $\sum |c_n|^2 = 0$. Hence (2) implies that $V(x) \equiv 0$.

4. *Solutions for $|x| \geqq R$.* It will be shown that an analogue of (I) holds for functions $V = V(x)$ which are not entire solutions of (1) but are solutions

on a set $|x| \geqq R > 0$. The function $V(ru)$ has, for a fixed r, a Fourier expansion of the form

$$r^{\frac{1}{2}(p-1)} V(ru) \sim (\tfrac{1}{2}\pi r)^{\frac{1}{2}} \sum_{n=0}^{\infty} i^n \sum_{j=1}^{N} \left(a_{nj} H_\mu^1(r) + b_{nj} H_\mu^2(r)\right) K_{nj}(u), \tag{19}$$

where $N = N(p, n)$ is the number of linearly independent spherical harmonics of order n; a_{nj} and b_{nj} are constants; H_μ^1 and H_μ^2 are the Hankel functions of order μ; and $K_{01}, K_{11} \ldots$ form a complete orthonormal sequence of spherical functions.

The form of the series in (19) merely reflects the fact that the Fourier coefficient

$$V_{nj}(r) = \int_U V(ru)\, \overline{K}_{nj}(u)\, du, \tag{20}$$

multiplied by $r^{\frac{1}{2}(p-2)}$ is a solution of the Bessel differential equation of index μ. The functions H_μ^1, H_μ^2 are solutions of the Bessel equation satisfying, for $k = 1, 2$,

$$(\tfrac{1}{2}\pi r)^{\frac{1}{2}} H_\mu^k(r) - \exp[(-1)^{k+1} i (r - \tfrac{1}{2}\mu\pi - \pi/4)] = o(1) \quad \text{as} \quad r \to \infty. \tag{21}$$

It is easily verified that the boundedness condition

$$r^{-1} \int_{R \leq |x| \leq r} |V(x)|^2\, dx \leqq C \qquad \text{for all } r \geqq R, \tag{22}$$

implies that

$$\sum_{n=0}^{\infty} \sum_{j=1}^{N} (|a_{nj}|^2 + |b_{nj}|^2) < \infty; \tag{23}$$

cf. the proof of (12).

Remark. It will follow from arguments below that, conversely, (19) and (23) imply (22) for some constant C.

Substitution of (21) into (19) gives formally the following approximation for $r^{\frac{1}{2}(p-1)} V(ru)$ for large r,

$$W(ru) = e^{ir} i^{\frac{1}{2}(1-p)} K^1(u) + e^{-ir} i^{\frac{1}{2}(p-1)} K^2(-u), \tag{24}$$

where

$$K^1(u) \sim \sum_{n=0}^{\infty} \sum_{j=1}^{N} a_{nj} K_{nj}(u), \qquad K^2(u) \sim \sum_{n=0}^{\infty} \sum_{j=1}^{N} b_{nj} K_{nj}(u). \tag{25}$$

It will be shown that the following analogue of (I) holds:

(III) *Let $V = V(x)$ be a solution of* (1) *for $|x| \geqq R$ satisfying* (22) *and let $W(x)$ be defined by* (19), (24), (25). *Then*

$$r^{-1} \int_R^r \|t^{\frac{1}{2}(p-1)} V(tu) - W(tu)\|^2\, dt \to 0 \quad \text{as} \quad r \to \infty, \tag{26}$$

and

$$r^{-1} \int_R^r \|t^{\frac{1}{2}(p-1)} V_r(tu) - W_r(tu)\|^2\, dt \to 0 \quad \text{as} \quad r \to \infty, \tag{27}$$

where $V_r = \partial V/\partial r$, $W_r = \partial W/\partial r$.

The functions H_μ^1, H_μ^2 are complex conjugates of each other and are given by $H_\mu^k = J_\mu + i(-1)^{k+1} Y_\mu$ in terms of the (real-valued) Bessel functions of the first and second kind. Thus, (19) can be written as

$$(28)\quad r^{\frac{1}{2}(p-1)} V(ru) \sim (\tfrac{1}{2}\pi r)^{\frac{1}{2}} \sum_{n=0}^{\infty} i^n \sum_{j=1}^{N} [(a_{nj} + b_{nj}) J_\mu(r) + i(a_{nj} - b_{nj}) Y_\mu(r)] K_{nj}(u).$$

Remark 1. The assertion (26) of (III) in the case that $a_{nj} - b_{nj} \equiv 0$ is essentially that of (I). The proof of (III) will show that, in the case $a_{nj} + b_{nj} \equiv 0$, (26) can be improved to

$$\| r^{\frac{1}{2}(p-1)} V(ru) - W(ru) \| \to 0 \quad \text{as} \quad r \to \infty.$$

Remark 2. The assertion (III) and the arguments leading to (II) give the following analogue of a theorem of RELLICH [4]: *Let $V(x)$ be as in* (III) *and let $V(ru) \to 0$ as $r \to \infty$ for almost all fixed u. Then $V(x) \equiv 0$.*

Remark 3. It follows from the Lemma of Section 6 below that (27) implies

$$(29)\quad r^{-1} \int_{R \le |x| \le r} |\nabla V|^2 \, dx \le \text{const.} \quad \text{for all} \quad r \ge R.$$

This fact will not be used in the proof of (III).

5. *Proof of* (26) *in* (III). The relation (28) and the proof of (I) make it clear that, in order to prove (26), it suffices to verify that both series

$$(30)\quad \begin{cases} \sum_{n=0}^{\infty} \sum_{j=1}^{N} |a_{nj} + b_{nj}|^2 (r - R)^{-1} \int_R^r t J_\mu^2(t) \, dt, \\ \sum_{n=0}^{\infty} \sum_{j=1}^{N} |a_{nj} - b_{nj}|^2 (r - R)^{-1} \int_R^r t Y_\mu^2(t) \, dt \end{cases}$$

are uniformly convergent for $R \le r < \infty$.

The uniform convergence of the first of these series is clear from (13) and (23).

The function $t(J_\mu^2(t) + Y_\mu^2(t))$ is non-increasing for $t > 0$ if $\mu \ge \frac{1}{2}$; cf. [5], p. 446. Thus $t Y_\mu^2(t) \le t(J_\mu^2(t) + Y_\mu^2(t)) \le R(J_\mu^2(R) + Y_\mu^2(R))$, so that the second series in (30) is majorized by

$$\sum_{n=0}^{\infty} \sum_{j=1}^{N} |a_{nj} - b_{nj}|^2 R(J_\mu^2(R) + Y_\mu^2(R)),$$

which is independent of r. Thus, it ony remains to verify the convergence of the last (numerical) series. A majorant for this series is furnished by the sum of the two series

$$2R \sum_{n=0}^{\infty} \sum_{j=1}^{N} |(a_{nj} + b_{nj}) J_\mu(R) + i(a_{nj} - b_{nj}) Y_\mu(R)|^2$$

$$2R \sum_{n=0}^{\infty} \sum_{j=1}^{N} (|a_{nj} + b_{nj}|^2 + |a_{nj} - b_{nj}|^2) J_\mu^2(R).$$

By (28), the first of these is convergent (and has the sum $\pi^{-1} \| R^{\frac{1}{2}(p-1)} V(Ru) \|^2$). By (23) and LOMMEL's inequality, $|J_\mu(r)| \le 1$ for $r, \mu \ge 0$ (cf. [5], p. 406), the second of the series is convergent. This proves (26).

6. *A lemma.* In order to avoid an interruption to the proof of (27) in (III), it will be convenient to isolate a simple lemma, which follows from a device of FRIEDRICHS [1]; cf. [6].

Lemma. *Let* $x=(x^1, \ldots, x^p)$, $p \geqq 1$. *Let* $q(x)$, $f(x)$ *be continuous functions for* $|x| \geqq R$ *and*

$$\operatorname{Re} q(x) \leqq M. \tag{31}$$

Let $X(x)$ *be a* C^2*-solution of*

$$\Delta X + q(x) X = f(x) \tag{32}$$

for $|x| \geqq R$, *and let* $T(a, b) = \{x : a \leqq |x| \leqq b\}$. *Then, for* $r > R+2$,

$$\int_{T(R+1, r-1)} |\nabla X|^2 \, dx \leqq 2 \int_{T(R, r)} [(M+8)|X|^2 + |fX|] \, dx. \tag{33}$$

This lemma will be used below only for the case that $p=1$. In this case, (33) can be interpreted as

$$\int_{R+1}^{r-1} |dX/dt|^2 \, dt \leqq 2 \int_{R}^{r} [(M+8)|X|^2 + |fX|] \, dt.$$

Proof. Let $\varepsilon(s)$ be a smooth function for $-\infty < s < \infty$ satisfying $\varepsilon(s) = 0$ for $s \leqq R$ and $s \geqq r$, $\varepsilon(s) = 1$ for $R+1 \leqq s \leqq r-1$, and $0 \leqq \varepsilon(s) \leqq 1$ and $|\varepsilon'(s)| \leqq 2$ for all s.

By GREEN'S formula, the relations

$$s^{p-1} \int_U \varepsilon^2(s) \, \overline{X}(su) \, X_r(su) \, du = 0 \quad \text{for} \quad s = r, R,$$

imply that

$$\int_{T(R, r)} \{\varepsilon^2 \overline{X} \, \Delta X + \varepsilon^2 |\nabla X|^2 + 2\varepsilon\varepsilon' \overline{X} X_r\} \, dx = 0,$$

where the argument of ε and ε' is $|x|$. Thus, by (32),

$$\int_{T(R, r)} \varepsilon^2 |\nabla X|^2 \, dx \overset{*}{=} \operatorname{Re} \int_{T(R, r)} [\varepsilon^2 (q|X|^2 - f\overline{X}) - 2\varepsilon\varepsilon' \overline{X} X_r] \, dx.$$

Since $|2\varepsilon\varepsilon' \overline{X} X_r| \leqq \frac{1}{2}\varepsilon^2 |X_r|^2 + 2\varepsilon'^2 |X|^2 \leqq \frac{1}{2}\varepsilon^2 |\nabla X|^2 + 8|X|^2$, it follows from (31) that

$$\tfrac{1}{2} \int_{T(R, r)} \varepsilon^2 |\nabla X|^2 \, dx \leqq \int_{T(R, r)} \left((M+8)|X|^2 + |fX|\right) dx.$$

Since $\varepsilon(s) = 1$ for $R+1 < s < r-1$, the inequality (33) follows.

7. *Proof of* (27) *in* (III). The relations (19) and (21) can be differentiated with respect to r to give

$$[r^{\frac{1}{2}(p-1)} V(ru)]_r \sim \sum_{n=0}^{\infty} i^n \sum_{j=1}^{N} \{a_{nj} [(\tfrac{1}{2}\pi r)^{\frac{1}{2}} H_\mu^1]_r + b_{nj} [(\tfrac{1}{2}\pi r)^{\frac{1}{2}} H_\mu^2]_r\} K_{nj} \tag{34}$$

and

$$[(\tfrac{1}{2}\pi r)^{\frac{1}{2}} H_\mu^k]_r + (-1)^k i \exp[(-1)^{k+1} i (r - \tfrac{1}{2}\mu\pi - \pi/4)] = o(1) \quad \text{as} \quad r \to \infty. \tag{35}$$

Thus the proofs of (I) and (26) in (III) make it clear that

$$r^{-1}\int_{R+1}^{r}\left\|\left[r^{\frac{1}{2}(p-1)}V(ru)\right]_r - W_r(ru)\right\|^2 dr \to 0 \quad \text{as} \quad r\to\infty \tag{36}$$

holds if the two series

$$\begin{cases} \sum\limits_{n=0}^{\infty}\sum\limits_{j=1}^{N}|a_{nj}+b_{nj}|^2 r^{-1}\int\limits_{R+1}^{r}\left(t^{\frac{1}{2}}J_\mu(t)\right)'^2 dt, \\ \sum\limits_{n=0}^{\infty}\sum\limits_{j=1}^{N}|a_{nj}-b_{nj}|^2 r^{-1}\int\limits_{R+1}^{r}\left(t^{\frac{1}{2}}Y_\mu(t)\right)'^2 dt \end{cases} \tag{37}$$

are uniformly convergent for $R-1\leqq r<\infty$.

The functions $X=t^{\frac{1}{2}}J_\mu(t)$ or $X=t^{\frac{1}{2}}Y_\mu(t)$ are solutions of the ordinary differential equation

$$X''+q(t)X=0, \quad \text{where} \quad q(t)=1+t^{-2}(1/4-\mu^2).$$

By (3), $n\geqq 1$ implies that $\mu>\frac{1}{2}$, so that $q(t)\leqq 1$. Hence, by the Lemma,

$$\int_{R+1}^{r}X'^2 dt \leqq 18\int_{R}^{r-1}X^2 dt.$$

Thus, the uniform convergence of the series (37) for $r>R+1$ follows from that of (30) for $r\geqq R$.

Hence (36) holds. Since

$$\left[r^{\frac{1}{2}(p-1)}V(ru)\right]_r = r^{\frac{1}{2}(p-1)}V_r(ru)+\tfrac{1}{2}(p-1)r^{-1}\left[r^{\frac{1}{2}(p-1)}V(ru)\right]$$

and since (22) implies that

$$r^{-1}\int_{R<|x|<r}|x|^{-2}|V(x)|^2 dx \to 0 \quad \text{as} \quad r\to\infty,$$

the relation (27) follows from (36). This proves (III).

References

[1] Friedrichs, K. O.: A theorem of Lichtenstein. Duke Math. J. **14**, 67–82 (1947). — [2] Magnus, W.: Fragen der Eindeutigkeit und des Verhaltens im Unendlichen für Lösungen von $\Delta u + k^2 u = 0$. Abh. Math. Sem. Hamburg **16**, 77–94 (1949). — [3] Müller, C.: Über die ganzen Lösungen der Wellengleichung. Math. Annalen **124**, 235–264 (1952). — [4] Rellich, F.: Über das asymptotische Verhalten der Lösungen von $\Delta u + \lambda u = 0$ in unendlichen Gebieten. Jahresber. Deutsche Math.-Vereinigung **53**, 57–65 (1943). — [5] Watson, G. N.: Theory of Bessel functions, 2nd ed. **1945**. — [6] Wintner, A.: On the momentum operator in wave mechanics. Phys. Rev. **71**, 547 to 549 (1947).

Dept. of Math., The Johns Hopkins University, Baltimore 18, Md. (U.S.A.)

(Eingegangen am 24. November 1958)

KOECHER, M., u. W. ROELCKE
Math. Zeitschr. 71, 258–267 (1959)

Diskontinuierliche und diskrete Gruppen von Isometrien metrischer Räume

Von
MAX KOECHER und WALTER ROELCKE

1. Sei X ein topologischer Raum und Ω eine Gruppe topologischer Selbstabbildungen von X. Eine Untergruppe Γ von Ω nennt man *diskontinuierlich*, wenn für jedes x aus X und jede Folge von verschiedenen α_n aus Γ die Folge $\alpha_n x$ in X keinen Häufungspunkt besitzt. Abgesehen von mannigfachen konkreten Beispielen wurde erstmals von C. L. SIEGEL[1]) der Zusammenhang der Begriffe diskontinuierlich und *diskret* (in einer geeigneten Topologie von Ω) untersucht. Der Satz von SIEGEL zeigt, daß beide Begriffe äquivalent sind, falls man Ω als topologische Gruppe vorgibt, in der das Einselement eine abzählbare Umgebungsbasis besitzt, und für X den Restklassenraum von Ω nach einer kompakten Untergruppe $\mathfrak{P}$ von Ω wählt, wobei $\mathfrak{P}$ keinen echten Normalteiler von Ω enthält. Ω operiert dann transitiv und effektiv auf dem homogenen Raum $X = \Omega/\mathfrak{P}$.

In der vorliegenden Note wird diese Fragestellung für metrische Räume X untersucht und dabei Ω als Gruppe der Isometrien von X gewählt. Während für metrische Räume die obige Definition von diskontinuierlich wohl die einzig vernünftige ist, liegt die Situation für die Definition von diskret nicht ganz so einfach. In Ω können nämlich verschiedene Topologien eingeführt werden und eine Untergruppe Γ von Ω wird man diskret (in der entsprechenden Topologie) nennen, wenn sie als Unterraum diskret ist. Das Ziel dieser Untersuchungen ist, geeignete und in natürlicher Weise gegebene Topologien von Ω zu finden, für die eine Untergruppe Γ möglichst genau dann diskret ist, wenn sie diskontinuierlich ist.

Da bei der hier gegebenen Beweisführung an mehreren Stellen die Lokal-Kompaktheit von X vorausgesetzt werden muß, wird diese stets vorausgesetzt und auf die teilweise möglichen Abschwächungen nicht eingegangen.

Für zusammenhängende metrische Räume kann ein Teil der Ergebnisse aus einer Arbeit von D. VAN DANTZIG und B. L. VAN DER WAERDEN[2]) entnommen werden, in der jedoch der Begriff diskontinuierlich nicht explizit auftritt. — Teile der Aussagen von Lemma 1 und Satz 1 sind in [4]) enthalten.

Wir danken Herrn A. DOLD sehr für einige wertvolle Hinweise.

2. Es sei X stets ein metrischer Raum mit der Metrik $|x, y|$ und X lokalkompakt in der Topologie, die durch die Metrik induziert wird. Für jede Teil-

[1]) SIEGEL, C. L.: Discontinuous Groups. Ann. Math. (2) **44**, 674–689, Lemma 6 (1943).
[2]) VAN DANTZIG, D., u. B. L. VAN DER WAERDEN: Über metrische homogene Räume. Abh. math. Sem. Hamburg **6**, 367–376 (1928).

menge A von X bezeichne wie üblich $\mathring{A}$ bzw. $\bar{A}$ den offenen Kern bzw. die abgeschlossene Hülle von A.

Eine Abbildung α von X auf sich heißt eine *Isometrie* von X, wenn die Metrik bei α invariant ist, d.h.

$$|\alpha x, \alpha y| = |x, y| \qquad \text{für alle } x \text{ und } y \text{ aus } X \tag{1}$$

gilt. Hierbei bedeutet αx das Bild von x bei der Abbildung α. Offenbar stellen die Isometrien topologische Selbstabbildungen von X dar und bilden eine Gruppe Ω. Das Einselement von Ω sei ι.

Für jedes Kompaktum K von X und $\alpha, \beta \in \Omega$ definieren wir

$$|\alpha, \beta|_K := \sup\{|\alpha x, \beta x|;\ x \in K\}. \tag{2}$$

Ersichtlich stellt $|\alpha, \beta|_K$ bei festem K auf dem direkten Produkt $\Omega \times \Omega$ eine nicht-negative Funktion dar, die der Dreiecksungleichung genügt und zugleich links-invariant ist, d.h. es gilt

$$|\gamma\alpha, \gamma\beta|_K = |\alpha, \beta|_K \qquad \text{für alle } \alpha, \beta, \gamma \in \Omega. \tag{3}$$

Weiter ist aber auch

$$|\alpha\gamma, \beta\gamma|_K = |\alpha, \beta|_{\gamma K} \qquad \text{für alle } \alpha, \beta, \gamma \in \Omega. \tag{4}$$

Aus dieser „Quasi-Metrik" leitet sich wie bei metrischen Räumen eine „K-Topologie" für Ω her: Für $\varepsilon > 0$ und $\alpha \in \Omega$ sei

$$U_\varepsilon(\alpha; K) := \{\gamma;\ \gamma \in \Omega, |\gamma, \alpha|_K < \varepsilon\}. \tag{5}$$

Wegen (3) und (4) gilt dann

$$U_\varepsilon(\alpha; K) = \alpha\, U_\varepsilon(\iota; K) = U_\varepsilon(\iota; \alpha K)\, \alpha.$$

Eine Teilmenge A von Ω nennen wir *K-offen*, wenn es zu jedem α aus A ein $\varepsilon > 0$ mit $U_\varepsilon(\alpha; K) \subset A$ gibt. Man sieht leicht, daß die $U_\varepsilon(\alpha; K)$ selbst, eine beliebige Vereinigung und jeder endliche Durchschnitt von K-offenen Mengen ebenfalls K-offen sind. Die K-offenen Mengen induzieren also in Ω eine Topologie, die *K-Topologie*. Die Menge der K-offenen Mengen von Ω werde mit $\mathfrak{T}_K$ bezeichnet. In dieser Topologie konvergiert eine Folge α_n aus Ω dann und nur dann gegen ein α aus Ω, falls $|\alpha, \alpha_n|_K \to 0$. Wir wollen dafür auch

$$\alpha_n \xrightarrow[K]{} \alpha$$

schreiben. Im allgemeinen ist in der K-Topologie keines der üblichen Trennungsaxiome erfüllt. Sind K und L Kompakta von X, und $K \subset L$, so ist das System der K-offenen Mengen offenbar enthalten im System der L-offenen Mengen von Ω, d.h. die K-Topologie ist gröber als die L-Topologie.

3. Wir bilden nun die „Vereinigungstopologie" aller K-Topologien. Wir bezeichnen diese neue Topologie als *ko-Topologie* von Ω. Sie ist feiner als jede K-Topologie. Eine Basis[3]) der *ko-offenen* — d.h. in der *ko*-Topologie

[3]) In dem Sinne, daß *ko*-offene Mengen — und nur diese — darstellbar sind als Vereinigung von Mengen aus der Basis.

offenen — Mengen von Ω besteht also aus sämtlichen K-offenen Mengen (genommen für alle Kompakta K von X) und aus den endlichen Durchschnitten dieser Mengen. Eine Basis für die *ko*-offenen Mengen von Ω bildet auch das Mengensystem

$$\mathfrak{S} := \{U_\varepsilon(\alpha; K);\ \varepsilon > 0,\ \alpha \in \Omega,\ K < X \text{ kompakt}\}.$$

Beim Nachweis beachte man für die Durchschnitte: Ist $A < \Omega$ K-offen und $B < \Omega$ L-offen (K, L Kompakta aus X), so sind A und B, und also auch $A \cap B$, $(K \cup L)$-offen und folglich darstellbar mit Hilfe der in $\mathfrak{S}$ enthaltenen Basis $\{U_\varepsilon(\alpha; K \cup L);\ \varepsilon > 0, \alpha \in \Omega\}$ für die $(K \cup L)$-Topologie.

In der *ko*-Topologie ist Ω ein Hausdorff-Raum: Sind $\alpha, \beta \in \Omega$, $\alpha \neq \beta$, so gibt es ein $x \in X$, so daß $\alpha x \neq \beta x$. Für das Kompaktum K von X, welches aus x allein besteht, ist $r := \frac{1}{2}|\alpha, \beta|_K = \frac{1}{2}|\alpha x, \beta x| > 0$. Die Umgebungen $U_r(\alpha; K)$, $U_r(\beta; K)$ von α, β in der K-Topologie sind zugleich Umgebungen in der *ko*-Topologie. Ihr Durchschnitt ist leer, denn für ein Element γ des Durchschnitts wäre

$$|\gamma, \alpha|_K + |\gamma, \beta|_K < r + r, \quad \text{also} \quad |\alpha, \beta|_K < 2r$$

im Widerspruch zur Bestimmung von r.

Konvergiert eine Folge $\alpha_n \in \Omega$ in der *ko*-Topologie gegen $\alpha \in \Omega$, dann schreiben wir auch

$$\alpha_n \xrightarrow[ko]{} \alpha.$$

LEMMA 1. *Für $\alpha_n, \alpha \in \Omega$ sind die folgenden drei Aussagen äquivalent:*

1) $\alpha_n \xrightarrow[ko]{} \alpha$,

2) $\alpha_n \xrightarrow[K]{} \alpha$ *für jedes Kompaktum K von X,*

3) $\alpha_n x \to \alpha x$ *für alle $x \in X$.*

BEWEIS. Da jede K-offene Menge auch *ko*-offen ist und die K-offenen Mengen eine Basis der *ko*-offenen Mengen bilden, sind 1) und 2) ersichtlich äquivalent. 3) folgt trivial aus 2). Es bleibt zu zeigen, daß 2) aus 3) folgt. Dazu ist zu zeigen, daß die Folge $\alpha_n x$ auf jedem Kompaktum K von X gleichmäßig konvergiert. Wäre das nicht richtig, dann gibt es ein $\varepsilon > 0$, eine Folge $x_n \in K$ und eine Teilfolge β_n von α_n mit

$$|\beta_n x_n, \alpha x_n| \geqq \varepsilon.$$

Hier können wir $x_n \to x \in K$ annehmen und erhalten

$$\begin{aligned} |\beta_n x_n, \alpha x_n| &\leqq |\beta_n x_n, \beta_n x| + |\beta_n x, \alpha x| + |\alpha x, \alpha x_n| \\ &\leqq 2|x_n, x| + |\beta_n x, \alpha x|. \end{aligned}$$

Wegen $|x_n, x| \to 0$ und $|\beta_n x, \alpha x| \to 0$ ist das ein Widerspruch.

Nach N. BOURBAKI[4]) ist die *ko*-Topologie identisch mit der wie folgt definierten „kompakt-offen"-Topologie. Bezeichnet K eine kompakte und W eine offene Teilmenge von X und setzt man $M(K, W) := \{\alpha;\ \alpha \in \Omega, \alpha K < W\}$, so bilden die endlichen Durchschnitte der Mengen $M(K, W)$ für variables K

[4]) BOURBAKI, N.: Topologie générale, Chap. X, § 2.

und W eine Basis für die kompakt-offen-Topologie. Man vergleiche auch S. B. MYERS[5]), Lemma 3.1.

4. Besitzt X eine abzählbare Topologie, so hat ι in der *ko*-Topologie eine abzählbare Umgebungsbasis. Denn dann gibt es eine aufsteigende Folge von Kompakta K_n $(n \geqq 1)$ von X mit

$$X = \bigcup_{n=1}^{\infty} \mathring{K}_n,$$

und da ι in jeder K-Topologie eine abzählbare Basis von Umgebungen besitzt und die Vereinigung dieser K-Basen eine *ko*-Basis für ι ist, folgt die Behauptung aus $\mathfrak{T}_K \subset \mathfrak{T}_{K_n}$ für $K \subset K_n$ und der Tatsache, daß jedes Kompaktum K in einem hinreichend späten K_n enthalten ist.

SATZ 1. 1) *Ω ist in der ko-Topologie eine topologische Gruppe.*

2) *Ω operiert als Transformationsgruppe im üblichen Sinne auf X, d.h. die Abbildung $\sigma: (\alpha, x) \to \alpha x$ $(\alpha \in \Omega, x \in X)$ von $\Omega \times X$ auf X ist ko-stetig.*

3) *Ist Λ eine beliebige Untergruppe von Ω, so ist die durch die ko-Topologie in Λ induzierte Topologie die gröbste unter allen Topologien, für die Λ als Transformationsgruppe auf X operiert.*

BEWEIS. 1) Da die Gruppeneigenschaft von Ω schon bekannt ist und da Ω hausdorffsch ist, bleibt zu zeigen, daß die Abbildung

$$\pi: (\alpha, \beta) \to \alpha^{-1}\beta \qquad (\alpha, \beta \in \Omega)$$

von $\Omega \times \Omega$ in Ω stetig ist. Da aber die *ko*-offenen Mengen als Vereinigung von K-offenen Mengen erhalten werden, genügt es, zu beweisen: Für jedes $\varepsilon > 0$ und jedes Kompaktum K von X enthält $U_\varepsilon(\alpha^{-1}\beta; K)$ das Bild πV der *ko*-Umgebung $V := U_{\varepsilon/2}(\alpha; \alpha^{-1}\beta K) \times U_{\varepsilon/2}(\beta; K)$ von (α, β) in $\Omega \times \Omega$. Sei dazu $(\alpha_1, \beta_1) \in V$, dann folgt

$$\begin{aligned} |\alpha_1^{-1}\beta_1, \alpha^{-1}\beta|_K &\leqq |\alpha_1^{-1}\beta_1, \alpha_1^{-1}\beta|_K + |\alpha_1^{-1}\beta, \alpha^{-1}\beta|_K \\ &\leqq |\beta_1, \beta|_K + |\alpha_1, \alpha|_{\alpha^{-1}\beta K} < \varepsilon, \end{aligned}$$

also die erste Behauptung.

2) Sei für den Augenblick $U_\varepsilon(x) := \{y;\ y \in X, |y, x| < \varepsilon\}$. Man sieht sofort, daß für beliebiges $U_\varepsilon(\alpha x)$ die offene Menge $V := U_{\varepsilon/2}(\alpha; \{x\}) \times U_\varepsilon(x) \subset \Omega \times X$ im Urbild $\sigma^{-1} U_\varepsilon(\alpha x)$ von $U_\varepsilon(\alpha x)$ enthalten ist, d.h. σ stetig ist.

3) Die von der *ko*-Topologie in Λ induzierte Topologie bezeichnen wir auch als *ko*-Topologie von Λ. Ist dann eine beliebige „$*$-Topologie" von Λ gegeben, für die Λ eine Transformationsgruppe von X darstellt, so genügt es zu zeigen: Zu jedem $\varepsilon > 0$ und Kompaktum K von X gibt es eine $*$-Umgebung U^* von ι in Λ, so daß $U^* \subset U_\varepsilon(\iota; K)$. (Man beachte, daß die $U_\varepsilon(\iota; K)$ eine Basis von *ko*-Umgebungen von ι in Ω bilden.) Nun stellt aber

$$f(\alpha) := \sup\{|\alpha x, x|;\ x \in K\}$$

für jedes Kompaktum K von X eine auf Λ $*$-stetige Funktion dar. Für jedes $\varepsilon > 0$ ist also die Menge $\{\alpha; \alpha \in \Lambda, f(\alpha) < \varepsilon\}$ eine $*$-Umgebung von ι. Offenbar ist sie in $U_\varepsilon(\iota; K)$ enthalten. Das vollendet den Beweis.

[5]) MYERS, S. B.: Equicontinuous sets of mappings. Ann. Math. **47**, 496–502 (1946).

Ist der lokal-kompakte metrische Raum X zusammenhängend, dann besitzt X bekanntlich[6]) eine abzählbare Topologie und nach VAN DANTZIG und VAN DER WAERDEN[7]) weiß man, daß Ω als lokal-kompakte topologische Gruppe aufgefaßt werden kann, so daß für jedes $x \in X$ und jedes Kompaktum K von X die Menge $\{\alpha; \alpha \in \Omega, \alpha x \in K\}$ kompakt ist. Insbesondere ist dann also für jedes $x \in X$ die „Rotationsgruppe" $\{\alpha; \alpha \in \Omega, \alpha x = x\}$ kompakt in Ω. Von der in Ω definierten Topologie (loc. cit., § 6) läßt sich zeigen, daß sie abzählbar und mit der *ko*-Topologie äquivalent ist.

5. Eine Untergruppe Γ von Ω wollen wir *K-diskret* bzw. *ko-diskret* nennen, wenn Γ in der durch die K- bzw. *ko*-Topologie von Ω induzierten Topologie diskret ist. Γ heißt weiter *lokal-effektiv*, wenn aus $\alpha x = x$ für alle x aus einer offenen Teilmenge $A \neq \emptyset$ von X stets $\alpha = \iota$ folgt. Γ möge schließlich *diskontinuierlich* (auf X) genannt werden, falls für jede Folge $\alpha_n \in \Gamma$ von verschiedenen Isometrien und jedes $x \in X$ die Folge $\alpha_n x$ divergiert. Da X lokal-kompakt ist, ist damit gleichwertig, daß jedes Kompaktum von X nur endlich viele $\alpha_n x$ enthält.

LEMMA 2. *Ist Γ K-diskret für alle Kompakta K, $\overset{\circ}{K} \neq \emptyset$, von X, dann ist Γ ko-diskret und lokal-effektiv.*

Das noch mehrfach benötigte Beispiel des „Dreibeins"[7]) zeigt, daß aus der *ko*-Diskretheit von Γ weder die K-Diskretheit für alle Kompakta folgt noch daß Γ lokal-effektiv ist. Trivialerweise ist aber auch die zweite Aussage des Lemmas nicht umkehrbar.

BEWEIS. Sei $\alpha \in \Gamma$. Nach Voraussetzung gibt es für ein Kompaktum K von X eine Umgebung U_K der K-Topologie von Ω mit $U_K \cap \Gamma = \{\alpha\}$. U_K ist auch eine *ko*-Umgebung von α, also ist Γ auch *ko*-diskret. Zum Beweis der zweiten Aussage sei $\emptyset \neq A = \overset{\circ}{A} \subset X$ und $\alpha \in \Gamma$ mit $\alpha x = x$ für alle $x \in A$ gegeben. Wir wählen ein Kompaktum K von X mit $\overset{\circ}{K} \neq \emptyset$, $K \subset A$. Es ist dann $|\alpha, \iota|_K = 0$. Da Γ K-diskret ist, gibt es $\varepsilon > 0$ mit $U_\varepsilon(\iota; K) \cap \Gamma = \{\iota\}$. Wegen $\alpha \in U_\varepsilon(\iota; K)$ folgt aber $\alpha = \iota$.

LEMMA 3. *Ist Γ diskontinuierlich, dann ist Γ ko-diskret.*

BEWEIS. (Nach einer Vereinfachung von Herrn DOLD.) Es sei $x \in X$ fest und $\varepsilon > 0$ so klein gewählt, daß die Kugel $K := \{y; y \in X, |y, x| \leqq \varepsilon\}$ kompakt ist. Dann gilt

$$U_\varepsilon(\iota; x) \cap \Gamma = \{\gamma; \gamma \in \Gamma, |\gamma, \iota|_x < \varepsilon\} \subset \{\gamma; \gamma \in \Gamma, \gamma x \in K\},$$

[6]) ALEXANDROFF, P.: Über die Metrisation der im Kleinen kompakten topologischen Räume. Math. Ann. **92**, 294–301 (1924). Einen einfacheren Beweis des hier erwähnten Satzes findet man auch bei A. PFLUGER, Theorie der Riemannschen Flächen, S. 22. Berlin-Göttingen-Heidelberg: Springer 1957.

[7]) Der als Dreibein bezeichnete metrische Raum besteht aus drei euklidischen Strecken der Länge 1, deren Anfangspunkte identifiziert sind. Als Entfernung zweier Punkte x, y des Dreibeins wird die Länge des kürzesten Streckenzuges erklärt, welcher x und y verbindet.

d.h. $U_\varepsilon(\iota; x) \cap \Gamma$ ist endlich. Da nun Γ gemäß Satz 1 in der *ko*-Topologie hausdorffsch ist, folgt die Behauptung.

LEMMA 4. *Ist Γ diskontinuierlich und lokal-effektiv, so ist Γ K-diskret für alle Kompakta K von X mit $\mathring{K} \neq \emptyset$.*

Am Beispiel des Dreibeins[7]) sieht man, daß Γ diskontinuierlich sein kann, ohne lokal-effektiv (bzw. K-diskret für alle K) zu sein.

BEWEIS. Indirekt: Es sei Γ diskontinuierlich und lokal-effektiv, K ein Kompaktum von X mit $\mathring{K} \neq \emptyset$ und Γ nicht K-diskret. Zu jeder natürlichen Zahl n gibt es dann ein $\alpha_n \neq \iota$ aus $\Gamma \cap U_{1,n}(\iota; K)$. Das bedeutet

$$|\alpha_n x, x| < \frac{1}{n} \qquad \text{für alle } x \in K \text{ und } n \geqq 1. \tag{6}$$

Da Γ diskontinuierlich und X lokal-kompakt ist, kann es unter den α_n nur endlich viele verschiedene geben. Es gibt also eine konstante Teilfolge $\beta_n = \alpha \neq \iota$ der α_n. Wegen (6) muß dann aber gelten

$$\alpha x = x \qquad \text{für alle } x \in K.$$

Wegen $\alpha \neq \iota$ und $\mathring{K} \neq \emptyset$ ist dies ein Widerspruch dazu, daß Γ lokal-effektiv vorausgesetzt war.

6. Das eigentliche Ziel der vorliegenden Note ist ein Beweis dafür, daß aus der Diskretheit (vgl. **5**) einer Untergruppe Γ von Ω folgt, daß Γ auch diskontinuierlich ist. Da wir in Ω verschiedene Topologien zur Verfügung haben, ist dies zuerst dahingehend zu präzisieren, welche Topologie gemeint ist. Da nach **3** die *ko*-Topologie feiner als jede K-Topologie ist, stellt die *ko*-Diskretheit eine schwächere Forderung an Γ als die K-Diskretheit. In der Tat hatten wir in Lemma 2 gesehen, daß Γ sicher *ko*-diskret ist, falls Γ als K-diskret für alle K vorausgesetzt wird, hier jedoch die Umkehrung nicht gilt. Danach wäre die Aussage, daß eine Untergruppe Γ von Ω schon dann diskontinuierlich ist, wenn sie nur *ko*-diskret ist — d.h. die Umkehrung von Lemma 3 — die bestmögliche. Dies ist jedoch nicht richtig, wie ein Gegenbeispiel zeigt[8]). In diesem Beispiel ist Γ sogar noch lokal-effektiv, was aus

---

[8]) Man wähle als Punktmenge X die Vereinigung von einer Zahlengeraden R und einer Kreislinie C. Die Metrik $|x, y|$ wird wie folgt erklärt.

$$|x, y| := \begin{cases} \min\{1, |x - y|\} & \text{für } x, y \in R; \\ 1 \text{ für } x \in R, \ y \in C; \\ \text{eukl. Abstand von } x, y \text{ für } x, y \in C. \end{cases}$$

Es sei dann α diejenige Selbstabbildung von X, die R in sich um 1 in der positiven Richtung verschiebt und gleichzeitig C um einen irrationalen Winkel in sich verdreht. α stellt eine Isometrie von X dar. Die von α erzeugte unendliche zyklische Gruppe Γ ist nicht diskontinuierlich, da die Bilder eines Punktes $x \in C$ unter Γ in C dicht liegen. Andererseits ist Γ *ko*-diskret. Denn wenn als Kompaktum K ein einziger Punkt aus R gewählt wird, ist Γ offenbar sogar K-diskret.

der Irrationalität des auftretenden Drehwinkels folgt. Das angegebene Beispiel zeigt ferner, daß Γ nicht K-diskret für alle K zu sein braucht, wenn Γ *ko*-diskret und lokal-effektiv ist (vgl. die Bemerkung im Anschluß an Lemma 2). Wir werden später aber sehen, daß Lemma 3 unter der zusätzlichen Voraussetzung des Zusammenhanges von X umkehrbar ist.

Ohne weitere Voraussetzungen können wir aber zeigen, daß Γ diskontinuierlich ist, falls wir die *ko*-Diskretheit durch die stärkere K-Diskretheit für alle K ersetzen. Das soll im nächsten Abschnitt ausgeführt werden; es sind dazu jedoch noch einige Vorbereitungen nötig.

LEMMA 5. *Es seien K, L Kompakta von X und α_n $(n \geqq 1)$ eine Folge aus Ω, so daß gilt $\alpha_n K \subset L$ $(n \geqq 1)$. Dann gibt es eine Teilfolge β_n der α_n, so daß die Folge $\beta_n x$ $(n \geqq 1)$ in X konvergiert, wie auch $x \in K$ gewählt wird.*

BEWEIS. Bekanntlich[9]) gibt es eine höchstens abzählbare Teilmenge $A = \{a_m; m \geqq 1\}$ von X, die in K dicht liegt. Da bei festem m die Folge $\alpha_n a_m$ $(n \geqq 1)$ im Kompaktum L enthalten ist, gibt es eine konvergente Teilfolge. Eine Teilfolge β_n der Folge α_n, die auf ganz A punktweise konvergiert, erhält man dann nach dem bekannten Diagonalverfahren. Wir zeigen jetzt, daß $\beta_n x$, $x \in K$, eine Cauchy-Folge ist. Sei also $x \in K$ und $\varepsilon > 0$ gegeben. Da A in K dicht liegt, gibt es $a_m \in A$ mit $|x, a_m| < \frac{1}{3}\varepsilon$. Wir hatten oben schon gesehen, daß die Folge $\beta_n a_m$ $(n \geqq 1)$ konvergiert, d.h. es gibt $N(\varepsilon)$ mit

$$|\beta_p a_m, \beta_q a_m| < \tfrac{1}{3}\varepsilon \qquad \text{für alle } p, q > N(\varepsilon).$$

Für alle $p, q > N(\varepsilon)$ ist jetzt aber

$$|\beta_p x, \beta_q x| \leqq |\beta_p x, \beta_p a_m| + |\beta_p a_m, \beta_q a_m| + |\beta_q a_m, \beta_q x| < \varepsilon$$

d.h. $\beta_n x$ ist eine Cauchy-Folge. Wegen $\beta_n x \in L$ folgt die Behauptung, denn L war kompakt.

LEMMA 6. *Ist Γ nicht diskontinuierlich, dann gibt es $x \in X$ und eine Folge von verschiedenen Isometrien $\alpha_n \in \Gamma$ mit $\alpha_n x \to x$.*

BEWEIS. Nach Voraussetzung gibt es $x, y \in X$ und verschiedene Isometrien $\beta_n \in \Gamma$ mit $\beta_n x \to y$. Ist hier $y = \alpha x$ mit einem geeigneten $\alpha \in \Gamma$, so ist nichts zu beweisen. Sei also $y \neq \alpha x$ für alle $\alpha \in \Gamma$. Setzen wir jetzt

$$\alpha_n := \beta_n^{-1} \beta_{n+1},$$

dann ist $\iota \neq \alpha_n \in \Gamma$ und man hat

$$|\alpha_n x, x| = |\beta_n x, \beta_{n+1} x| \to |y, y| = 0,$$

d.h. $\alpha_n x \to x$. Es ist noch zu zeigen, daß es unter den α_n eine unendliche Folge von verschiedenen Isometrien gibt. Wäre das nicht richtig, so könnte wegen $\alpha_n x \to x$ von vornherein $\alpha_n x = x$ für alle n angenommen werden. Aus $\beta_{n+1} = \beta_n \alpha_n$ folgt dann aber $\beta_{n+1} x = \beta_n x$ und also $\beta_1 x = y$ im Widerspruch zur Annahme.

[9]) ALEXANDROFF, P., u. H. HOPF: Topologie, Kap. II, § 1.4. Berlin: Springer 1935.

LEMMA 7. *Ist Γ nicht diskontinuierlich, dann gibt es ein Kompaktum K, $\mathring{K} \neq \emptyset$, von X und eine Folge $\beta_n \in \Gamma$, $\beta_n \neq \iota$, mit $\beta_n \xrightarrow[K]{} \iota$.*

BEWEIS. Wir bestimmen $x \in X$ und $\alpha_n \in \Gamma$ nach Lemma 6. Da X lokalkompakt vorausgesetzt war, sind

$$L := \{y; |y, x| \leqq 2\varepsilon\} \quad \text{und} \quad K := \{y; |y, x| \leqq \varepsilon\}$$

für hinreichend kleines $\varepsilon > 0$ kompakt und $\mathring{K} \neq \emptyset$. Für jedes y aus K hat man aber

$$|\alpha_n y, x| \leqq |\alpha_n y, \alpha_n x| + |\alpha_n x, x| \leqq \varepsilon + |\alpha_n x, x| \leqq 2\varepsilon$$

für alle $n \geqq n_0$. Das heißt, daß für $n \geqq n_0$ alle Bilder $\alpha_n K$ von K in L enthalten sind. Nach Lemma 5 gibt es jetzt eine Teilfolge γ_n von α_n, für die die Folge $\gamma_n y$ für jedes y aus K konvergiert. Da die α_n nach Lemma 6 sämtlich verschieden sind, sind auch die γ_n sämtlich verschieden. Bei $\beta_n := \gamma_n^{-1}\gamma_{n+1}$ ist dann

$$\beta_n \neq \iota \quad \text{und} \quad |\beta_n y, y| = |\gamma_{n+1} y, \gamma_n y| \to 0 \qquad (y \in \mathring{K}).$$

Da K kompakt war, erhält man hieraus $|\beta_n, \iota|_K \to 0$ und das war die Behauptung.

7. Wir kommen jetzt zum Beweis der bereits angekündigten Sätze. Setzen wir voraus, daß eine Untergruppe Γ von Ω K-diskret ist für alle Kompakta K von X. Wäre Γ nicht diskontinuierlich, dann können wir nach Lemma 7 ein Kompaktum L von X, $\mathring{L} \neq \emptyset$ und eine Folge $\beta_n \in \Gamma$, $\beta_n \neq \iota$, finden, für die $|\beta_n, \iota|_L \to 0$ gilt. Da Γ auch L-diskret ist, gibt es ein $\varepsilon > 0$ mit $U_\varepsilon(\iota; L) \cap \Gamma = \{\iota\}$, d.h. aus $|\beta_n; \iota|_L < \varepsilon$ folgt $\beta_n = \iota$. Da aber alle β_n von der Identität verschieden waren, hat man einen Widerspruch, d.h. es gilt der folgende

SATZ 2. *Ist eine Untergruppe Γ von Ω K-diskret für alle Kompakta K von X, $\mathring{K} \neq \emptyset$, dann ist Γ auch diskontinuierlich.*

SATZ 3. *Sei X zusammenhängend und Γ eine ko-diskrete Untergruppe von Ω. Dann ist Γ auf X diskontinuierlich.*

BEWEIS. Wir stützen uns auf

LEMMA 8. *Ist X zusammenhängend, dann besitzt jede Folge α_n $(n \geqq 1)$ aus Ω, durch die ein Punkt $x \in X$ in eine konvergente Folge übergeführt wird, eine in der ko-Topologie von Ω konvergente Teilfolge.*

Eine wegen Lemma 1 gleichwertige Aussage ist in [2]), § 10—12, unter der wegen [6]) überflüssigen zusätzlichen Voraussetzung der Separabilität von X bewiesen. — Wäre Γ nicht diskontinuierlich, so gäbe es nun gerade eine solche Folge $\alpha_n \in \Gamma \subset \Omega$, und zwar mit sämtlich verschiedenen α_n. Die Existenz einer *ko*-konvergenten Teilfolge β_n der α_n steht dann im Widerspruch zur *ko*-Diskretheit von Γ: Man betrachte die Folge $\beta_n^{-1}\beta_{n+1}$.

Für zusammenhängendes X folgt Satz 2 wegen Lemma 2 offenbar auch aus Satz 3. — Die Stärke der Forderung, daß X zusammenhängend ist, zeigt sich auch in ihrer Auswirkung auf die bisher benutzten Topologien:

LEMMA 9. *Es sei X zusammenhängend, Λ eine lokal-effektive, ko-abgeschlossene Untergruppe von Ω. Dann sind für jedes Kompaktum $K \subset X$, $\mathring{K} \neq \emptyset$, die ko-Topologie und die K-Topologie des Relativraums Λ äquivalent.*

BEWEIS. Es genügt zu zeigen, daß jede (im Relativraum Λ) L-offene Teilmenge von Λ (L Kompaktum aus X) auch K-offen ist. Das ist genau dann der Fall, wenn $U_\varepsilon(\iota; L) \cap \Lambda$ ($\varepsilon > 0$) stets $U_\delta(\iota; K) \cap \Lambda$ für hinreichend kleines $\delta > 0$ enthält. Andernfalls gibt es ein $\varepsilon > 0$ und Folgen $\alpha_n \in \Lambda$ und $x_n \in L$ mit

$$|\alpha_n x, x| < \frac{1}{n} \quad \text{für alle } x \in K \tag{7}$$

und

$$|\alpha_n x_n, x_n| \geqq \varepsilon, \quad x_n \to x_0 \in L \tag{8}$$

für alle n. Gemäß Lemma 8 gibt es dann ein $\alpha \in \Omega$ und eine Teilfolge β_n der α_n, so daß $\beta_n x \to \alpha x$ für alle $x \in X$. Wegen der Abgeschlossenheit von Λ ist sogar $\alpha \in \Lambda$. Aus (7) folgt $\alpha x = x$ für alle $x \in K$ und daher $\alpha = \iota$, da Λ lokal-effektiv ist. Andererseits ist wegen (8) $|\alpha x_0, x_0| \geqq \varepsilon$, und man hat einen Widerspruch.

8. Die Nützlichkeit des in **5** eingeführten Begriffes „lokal-effektiv" erweist sich auch in dem folgenden

SATZ 4. *Es sei X ein — nicht notwendig lokal-kompakter — metrischer Raum und Γ sei eine auf X diskontinuierliche, lokal-effektive Gruppe von Isometrien von X. Dann bilden die Nichtfixpunkte von Γ*[10]) *eine in X dichte, offene Menge.*

Das Beispiel des Dreibeins[7]) zeigt, daß hier die Voraussetzung, daß Γ lokal-effektiv ist, nicht fortgelassen werden kann.

BEWEIS. 1) Lägen die Nichtfixpunkte nicht dicht in X, so gäbe es ein $x_0 \in X$ und eine Umgebung U von x_0 in X, die aus lauter Fixpunkten besteht. Zu jedem $x \in U$ wählen wir ein festes $\alpha_x \in \Gamma$, $\alpha_x \neq \iota$ mit $\alpha_x x = x$. Es gibt dann eine Umgebung $V \subset U$ von x_0, so daß die α_x für $x \in V$ nur endlich viele verschiedene Elemente aus Γ darstellen. Andernfalls gäbe es nämlich eine Folge x_n ($n \geqq 1$) in X mit $\lim_{n \to \infty} x_n = x_0$, so daß die Elemente $\alpha_n := \alpha_{x_n}$ sämtlich verschieden sind, und man erhielte dann wegen $x_n = \alpha_n x_n$ auf Grund der Dreiecksungleichung, und da die α_n Isometrien sind,

$$|x_0, \alpha_n x_0| \leqq |x_0, x_n| + |\alpha_n x_n, \alpha_n x_0| = 2|x_0, x_n| \to 0, \tag{9}$$

im Widerspruch zur Diskontinuierlichkeit von Γ.

Ist also V von der genannten Art gewählt, so seien, in neuer Bezeichnung, $\alpha_1, \ldots, \alpha_n$ die endlich vielen verschiedenen Isometrien unter den α_x, $x \in V$, also

$$\alpha_\nu \neq \iota \qquad (1 \leqq \nu \leqq n). \tag{10}$$

Die Nichtfixpunkte von α_ν ($1 \leqq \nu \leqq n$) bilden eine offene, in X dichte Menge A_ν. Denn A_ν ist offen als Komplementärmenge der abgeschlossenen Menge der

[10]) $x \in X$ heißt *Fixpunkt* von Γ, wenn $\alpha x = x$ mit einem $\alpha \in \Gamma$, $\alpha \neq \iota$ gilt; andernfalls heißt x Nichtfixpunkt von Γ.

Fixpunkte von α_ν, und wäre A_ν nicht dicht in X, so gäbe es eine nichtleere offene Menge in X, die nur aus Fixpunkten von α_ν besteht, und da Γ lokaleffektiv ist, folgte $\alpha_\nu = \iota$ im Widerspruch zu (10).

Nacheinander ergibt sich nun, daß die Mengen $V \cap \bigcap_{\mu=1}^{m} A_\mu$ $(1 \leqq m \leqq n)$ nichtleer (und offen) sind. Für $m = n$ ist das aber ein Widerspruch dazu, daß jedes $x \in V$ Fixpunkt von mindestens einer Transformation α_μ war. Damit ist bewiesen, daß die Nichtfixpunkte in X dicht liegen.

2) Die Nichtfixpunkte bilden eine offene Menge. Andernfalls gäbe es einen Nichtfixpunkt $x_0 \in X$ und eine Folge von Fixpunkten $x_n \in X$ mit $\lim x_n = x_0$ und $\alpha_n x_n = x_n$ für gewisse $\alpha_n \in \Gamma$, $\alpha_n \neq \iota$ $(n \geqq 1)$. Dann wäre aber wieder (9) erfüllt, wobei die rechte Seite eine Nullfolge darstellt. Da Γ diskontinuierlich ist, bedeutet das $\alpha_n x_0 = x_0$ $(\alpha_n \neq \iota)$ für fast alle n im Widerspruch zur Voraussetzung über x_0. Damit ist alles bewiesen.

Mathematisches Institut der Universität Münster

(Eingegangen am 10. August 1958)

NICKEL, K.
Math. Zeitschr. 71, 268–282 (1959)

Fehlerabschätzungen bei parabolischen Differentialgleichungen [1])

Von
KARL NICKEL

In dieser Note sollen Aussagen gemacht werden über Lösungen $z(x, y)$ der expliziten partiellen parabolischen Differentialgleichung (DGl)

$$z_x = f(x, y, z, z_y, z_{yy}) \tag{1}$$

unter gewissen Randbedingungen. Das Ziel ist die Gewinnung von Abschätzungen für einen allgemeinen Abstand $\varrho(x, y, z, v)$ einer Näherungslösung v einer zu (1) „benachbarten" DGl von den Lösungen z von (1). Fallen (1) und die benachbarte DGl zusammen, so ergeben sich auch Einzigkeitskriterien. Die Übertragung der gewonnenen Ergebnisse auf Systeme von DGlen der Gestalt (1) ist fast wörtlich möglich, soll jedoch zur Vereinfachung der Schreibweise hier nicht durchgeführt werden. Die ersten Abschätzungssätze über Lösungen von (1) wurden unabhängig voneinander von M. NAGUMO [1] und H. WESTPHAL [2] angegeben. Verallgemeinerungen finden sich bei M. NAGUMO-S. SIMODA [3], G. PRODI [4], R. NARASIMHAN [5], J. SZARSKI [6], L. COLLATZ [7], W. MLAK [8], K. NICKEL [9] und L. COLLATZ [12]. Da in den zitierten Arbeiten Beispiele zu Fehlerabschätzungen (insbesondere für *nicht*lineare DGlen der Gestalt (1)) kaum, bzw. überhaupt nicht gebracht werden, soll im folgenden besonderer Wert darauf gelegt werden, die Anwendbarkeit des zu beweisenden Abschätzungssatzes 2 an Hand von Beispielen zu zeigen.

Voraussetzungen

Es sei G eine beschränkte [2]), offene Punktmenge des $(m+1)$-dimensionalen euklidischen (x, y)-Raumes der Veränderlichen x und $y=(y_1, \ldots, y_m)$. Das kleinste offene Intervall, das die Projektion von G auf die x-Achse enthält, sei die Strecke $x_0 < x < x_1$. Für die Randpunkte von G werden folgende Bezeichnungen eingeführt: Γ_1 sei die Menge derjenigen Randpunkte $(\bar{x}, \bar{y})$, zu denen es eine Halbumgebung $0 \leqq \bar{x} - x < \delta$, $|\bar{y} - y| < \delta$ gibt [3]), die ganz in G liegt. Γ enthalte die restlichen Randpunkte von G (vgl. Fig. 1 und M. NAGUMO-S. SIMODA [3]). Γ zerfalle in zwei Teilmengen $\Gamma = \Gamma^* + \Gamma_r$. Von jedem Punkt

[1]) Über die folgenden Ergebnisse wurde schon kurz auf dem Internationalen Mathematiker-Kongreß 1958 in Edinburgh berichtet.

[2]) Zur Übertragung der folgenden Ergebnisse auf unbeschränkte Gebiete G vgl. man K. NICKEL [9], S. 6/7 und L. COLLATZ [12].

[3]) Es sei wie üblich $|y|^2 = \sum_{\mu=1}^{m} y_\mu^2$.

$P=(x', y') \in \Gamma_\nu$ aus gebe es eine auf der Hyperebene $x=x'$ liegende Halbgerade ν[4]), die ein Geradenstück mit P als einem Endpunkt mit $G+\Gamma_1$ gemeinsam hat[5]). Die Menge Γ_ν kann leer sein, dagegen enthält Γ^* mindestens die Randpunkte von G mit $x=x_0$. Bezeichnet man die abgeschlossene Hülle von G mit B, so gilt $B=G+\Gamma+\Gamma_1$.

Ist die quadratische $m \times m$-Matrix $A \equiv (a_{ik})$ positiv semidefinit, d.h. ist $\sum_{i,k=1}^{m} a_{ik} x_i x_k \geqq 0$ für alle $x_i (i=1, \ldots, m)$, dann soll kurz $A \geqq 0$ geschrieben werden. Statt $-A \equiv (-a_{ik}) \geqq 0$ werde auch die Abkürzung $A \leqq 0$, statt $A-B \equiv (a_{ik}-b_{ik}) \geqq 0$ ($B=(b_{ik})$ sei eine $m \times m$-Matrix) die Schreibweise $A \geqq B$ gebraucht.

Es seien die Funktionen $f(x, y_1, \ldots, y_m, r, s_1, \ldots, s_m, t_{11}, t_{12}, \ldots, t_{mm}) = f(x, y, r, s, t)$ und $g(x, y, r, s, t)$ definiert für $(x, y) \in G+\Gamma_1$ und beliebige Werte r, $s=(s_1, \ldots, s_m)$, $t=(t_{11}, t_{12}, \ldots, t_{mm})$; und f sei dort schwach monoton wachsend bezüglich t. Das soll heißen: Es ist $f(x, y, r, s, \bar{t}) \geqq f(x, y, r, s, t)$, wenn die Matrix $\bar{t}-t=\{\bar{t}_{ik}-t_{ik}\} \geqq 0$, d.h. positiv semidefinit ist. Die Funktionen $h(x, y_1, \ldots, y_m, r) = h(x, y, r)$ und $k(x, y, r)$ seien definiert für $(x, y) \in \Gamma_\nu$ und beliebige Werte r. Die Funktion $z(x, y_1, \ldots, y_m) = z(x, y)$ soll eine „zulässige Funktion" heißen, wenn gilt: Es sei $z \in C_0$[6]) auf B[7]), in $G+\Gamma_1$ mögen die partiellen Ableitungen $z_i = \dfrac{\partial z}{\partial y_i}$, $z_{ik} = \dfrac{\partial^2 z}{\partial y_i \partial y_k}$ und $z_x = \dfrac{\partial z}{\partial x}$[8]) existieren mit $z_i \in C_0$, $z_{ik} \in C_0$ und auf Γ_ν existiere noch die Ableitung z_ν von z in Richtung ν. Zur Abkürzung werde $z_y=(z_1, \ldots, z_m)$, $z_{yy}=\{z_{ik}\}$ geschrieben. Die Gesamtheit der zulässigen Funktionen bilde die Klasse Z. Schließlich sei noch die Funktion $\zeta(x, y)$ definiert und $\in C_0$[7]) für $(x, y) \in \Gamma^*$.

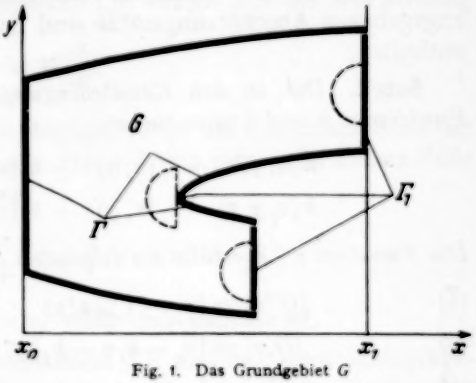


Fig. 1. Das Grundgebiet G

Mit diesen Bezeichnungen und Funktionen werde das folgende Anfangs-Randwertproblem AR definiert:

$$\text{(2)} \qquad \text{AR:} \begin{cases} Tz \equiv z_x - f(x, y, z, z_y, z_{yy}) = 0 & \text{in } G+\Gamma_1, \\ R^* z \equiv z - \zeta = 0 & \text{auf } \Gamma^*, \\ R_\nu z \equiv z_\nu - h(x, y, z) = 0 & \text{auf } \Gamma_\nu. \end{cases}$$

[4]) ν tritt hier an die Stelle der inneren Normale.

[5]) Die folgenden Aussagen lassen sich auch noch auf den allgemeineren Fall übertragen, daß ν ganz im Halbraum $x \leqq x'$ liegt.

[6]) Die Klasse C_k umfasse alle Funktionen, die in einem Gebiet stetig und mit stetigen partiellen Ableitungen bis zur Ordnung k versehen sind. Werden Randpunkte des Gebiets mitbetrachtet (hier die Punkte auf Γ_1), dann seien die Funktionen samt ihren k ersten Ableitungen stetig auf diese Punkte hinauf fortsetzbar.

[7]) Diese Forderung kann auf Γ^* noch abgemildert werden, man vgl. M. Nagumo - S. Simoda [3], S. 537, 2° und K. Nickel [9], S. 9, 2. Zusatz.

[8]) Es genügt schon, wenn nur die linke obere partielle Ableitung von z nach x existiert.

wobei als Lösungen $z(x, y)$ nur Funktionen $z \in Z$ zugelassen sind. Zu AR gehöre das „benachbarte" Anfangs-Randwertproblem AR':

$$\text{(3)} \qquad \text{AR}': \begin{cases} Sz \equiv z_x - g(x, y, z, z_y, z_{yy}) = 0 & \text{in } G + \Gamma_1, \\ Q^* z \equiv R^* z = z - \zeta = 0 & \text{auf } \Gamma^*, \\ Q_\nu z \equiv z_\nu - h(x, y, z) = 0 & \text{auf } \Gamma_\nu. \end{cases}$$

Ist Γ_ν leer, so liegt in beiden Fällen das erste Anfangs-Randwertproblem vor. Zwei Beispiele für allgemeine nichtlineare Randbedingungen der Gestalt $R_\nu z = 0$ findet man bei L. COLLATZ [7], S. 2 Fußnote, sowie bei K. NICKEL [9], S. 20, Fußnote 17.

Es sei nun $v \in Z$ eine Näherungslösung von AR', es soll der „Abstand" $z - v$ von v zu allen Lösungen $z \in Z$ von AR abgeschätzt werden. Die bisher angegebenen Abschätzungssätze sind im wesentlichen in dem folgenden Satz enthalten:

Satz 1. *Die in den Randbedingungen $R_\nu z = 0$ und $Q_\nu z = 0$ auftretenden Funktionen h und k seien linear:*

$$h(x, y, r) = h_1(x, y)\, r + h_2(x, y) \quad \text{mit} \quad h_1(x, y) > 0,$$
$$k(x, y, r) = k_1(x, y)\, r + k_2(x, y).$$

Die Funktion $v \in Z$ erfülle die folgenden Abschätzungen:

$$\text{(4)} \qquad |Q^* v| \equiv |v - \zeta| \leqq \varepsilon(x) \qquad \text{auf } \Gamma^*,$$

$$\text{(5)} \qquad |Q_\nu v| \equiv |v_\nu - k_1 v - k_2| \leqq \vartheta(x, y) \qquad \text{auf } \Gamma_\nu,$$

$$\text{(6)} \qquad |Sv| \equiv |v_x - g(x, y, v, v_y, v_{yy})| \leqq \delta(x) \qquad \text{in } G + \Gamma_1$$

und

$$\text{(7)} \quad f(x, y, s, v_y, v_{yy}) - g(x, y, t, v_y, v_{yy}) \leqq \omega(x, s - t) \quad \text{für alle } s \geqq t \text{ in } G + \Gamma_1.$$

Darin sei $\varepsilon(x)$ definiert in $\langle x_0, x_1 \rangle$, $\vartheta(x, y)$ definiert auf Γ_ν, $\delta(x) \in C_0$ in $\langle x_0, x_1 \rangle$ und $\omega(x, r) \in C_0$ für $x \in \langle x_0, x_1 \rangle$, r beliebig. Bezeichnet man mit $\sigma(x)$ das Maximalintegral der gewöhnlichen expliziten DGl erster Ordnung

$$\text{(8)} \qquad z' = \omega(x, z) + \delta(x) \quad \text{mit} \quad z(x_0) = \varepsilon(x_0)$$

in $\langle x_0, x_1 \rangle$[9]*, und genügt $\sigma(x)$ den Abschätzungen*[10]

$$\text{(9)} \qquad \sigma(x) \geqq \varepsilon(x) \quad \text{in} \quad \langle x_0, x_1 \rangle,$$

$$\text{(10)} \qquad \begin{cases} h_1(x, y)\, \sigma(x) \geqq \vartheta(x, y) + |v(x, y)\,[k_1(x, y) - h_1(x, y)] + \\ \qquad\qquad + k_2(x, y) - h_2(x, y)| \qquad \text{auf } \Gamma_\nu, \end{cases}$$

dann gilt für jede Lösung $z \in Z$ des Anfangs-Randwertproblems AR auf ganz B die Abschätzung

$$\text{(11)} \qquad |z - v| \leqq \sigma(x).$$

[9]) $\sigma(x) = \infty$ ist zugelassen.

[10]) Das ist stets durch passende Vergrößerung von $\varepsilon(x_0)$, $\omega(x, r)$ oder/und $\delta(x)$ zu erreichen.

Da dieser Satz 1 im wesentlichen aus dem nachstehenden allgemeineren Satz 2 folgt, soll sein Beweis nicht gebracht werden. Es werde jedoch auf die Ungleichung (7) hingewiesen, die offenbar die wichtigste Bedingung dieses Satzes darstellt. Es scheint bisher noch nicht beachtet worden zu sein, daß zur *beidseitigen* Abschätzung von $z-v$ nach (11) in Ungleichung (7) schon eine *einseitige* Bindung genügt, daß also die linke Seite von (7) nicht in Absolutstrichen zu stehen braucht, wenn nur die leicht nachzuprüfende Forderung (9) verlangt wird[11]). Damit sind nunmehr auch negative Funktionen $\omega(x, r)$ zugelassen, wodurch die Schranke $\sigma(x)$ in (11) u.U. erheblich verkleinert werden kann. Das zeigt das folgende einfache

Beispiel 1. Es sei Γ_r leer, $\varepsilon(x) \leqq \varepsilon$, $\delta(x) \leqq \delta$, $\omega(x, r) = -Kr$ mit $K > 0$ Ist dann ε so klein, daß $\varepsilon K \leqq \delta$ gilt, dann gilt auf ganz B die Abschätzung

$$|z - v| \leqq \frac{\delta}{K} \tag{12}$$

für jede Lösung $z \in Z$ von AR. Würde dagegen in (7) auf der linken Seite der absolute Betrag genommen, so wäre $\omega(x, r) = +Kr$ zu setzen. In diesem Falle ergäbe sich statt (12) die Abschätzung

$$|z - v| \leqq \left(\varepsilon(x_0) + \frac{\delta}{K}\right) \exp K(x - x_0) - \frac{\delta}{K}\,, \tag{13}$$

die offenbar für große Werte von $x - x_0$ völlig unbrauchbar ist.

Bei gegebenem Anfangs-Randwertproblem AR und fest gewähltem Näherungsproblem AR′ sind die in dem Abschätzungssatz auftretenden Funktionen $\varepsilon(x)$, $\vartheta(x, y)$, $\delta(x)$, $\omega(x, r)$ und $\sigma(x)$ im allgemeinen durch die verwendete Näherungslösung v vollständig bestimmt. Die sich ergebende Abschätzung (11) braucht jedoch keineswegs gut zu sein, da einerseits durch (4) bis (8) nur wenige Eigenschaften des Anfangs-Randwertproblems AR erfaßt werden, so daß solch eine Abschätzung die ungünstigsten Fälle mitberücksichtigen muß und da andererseits durch die eindimensionale Funktion $\sigma(x)$ die Abhängigkeit der Differenzfunktion $z - v$ von $m+1$ Veränderlichen im allgemeinen nur schlecht wiedergegeben werden kann. Der an sich naheliegende Übergang zu einer $m+1$-dimensionalen Schranke $\sigma(x, y)$ würde die zweite Schwierigkeit zwar beheben, doch wäre dann statt der gewöhnlichen DGl (8) eine entsprechende partielle parabolische DGl zu lösen. Um diesen Schritt zu vermeiden, soll im folgenden statt des Abstands $z - v$ der Lösungen z des Anfangs-Randwertproblems ein allgemeinerer Abstand $\varrho(x, y, z, v)$ behandelt und wieder durch die Lösung $\sigma(x)$ einer gewöhnlichen DGl abgeschätzt werden. Da ϱ weitgehend frei gewählt werden kann, ist es häufig möglich, zu günstigen Abschätzungen zu gelangen, indem man bekannte Eigenschaften der gegebenen Funktionen f, g, h, k, ζ, des Gebiets G oder der Lösungen z und v bei der Wahl von ϱ ausnutzt.

[11]) Bei der Abschätzung gewöhnlicher DGlen ist die entsprechende Tatsache seit langem bekannt. Man vgl. dazu etwa M. MÜLLER [*10*], S. 23, Satz 11 Formel (38).

Ein Beispiel — der Einfachheit halber aus dem Gebiet der gewöhnlichen DGlen — soll dies verdeutlichen: Die Lösungen der DGl

$$z'(x) = z^2(x) \quad \text{mit} \quad z(\xi) = \eta \tag{14}$$

sind die Funktionen $z(x) = \dfrac{1}{\frac{1}{\eta} + \xi - x}$[12]. Wegen der Singularität bei $\frac{1}{\eta} + \xi$ ist eine Abschätzung etwa zweier Lösungen z und v von (14) in der Form $|z - v| \leqq \sigma(x)$ wie in (11) mit einer überall stetigen Funktion $\sigma(x)$ nicht möglich (wohl natürlich für beschränkte x-Intervalle). Jedoch sind $1/z$ und $1/v$ auf der ganzen x-Achse regulär, eine Abschätzung $\frac{1}{z} - \frac{1}{v} \leqq \sigma(x)$ mit $\sigma \in C_0$ ist daher für alle x leicht möglich. Der Abstandsbegriff $\varrho(z, v) = \frac{1}{z} - \frac{1}{v}$ ist daher für die DGl (14) „angemessen".

Satz 2. *Die Funktion $v \in Z$ erfülle mit der Abstandsfunktion $\varrho(x, y, s, t)$ die folgenden Abschätzungen für alle $s \in \langle \alpha(x, y), \beta(x, y) \rangle$:*

$$|\varrho(x, y, \zeta, v)| \leqq \varepsilon(x) \qquad \text{auf } \Gamma^*, \tag{15}$$

$$|\varrho_t(x, y, s, v)\, Q_* v| \leqq \vartheta(x, y) \qquad \text{auf } \Gamma_*, \tag{16}$$

$$|\varrho_t(x, y, s, v)\, S v| \leqq \delta(x) \qquad \text{in } G + \Gamma_1 \tag{17}$$

und

$$\begin{cases} \operatorname{sign}(s - v)\,[\varrho_x(x, y, s, v) + \varrho_s(x, y, s, v)\, f\,(x, y, s, A(s), B(s)) + \\ + \varrho_t(x, y, s, v)\, g(x, y, v, v_y, v_{yy})] \leqq \omega\,(x, |\varrho(x, y, s, v)|) \end{cases} \qquad \text{in } G + \Gamma_1. \tag{18}$$

Dabei sei in (18) *zur Abkürzung gesetzt:*

$$A(s) = -\frac{1}{\varrho_s}\,[\varrho_t v_y + \varrho_y], \tag{19}$$

$$\begin{cases} B(s) = -\dfrac{1}{\varrho_s}\Big[\varrho_t v_{yy} + 2 v_y \Big(\varrho_{yt} - \dfrac{\varrho_{ys}}{\varrho_s}\varrho_t - \dfrac{\varrho_{st}}{\varrho_s}\varrho_y + \dfrac{\varrho_{ss}}{\varrho_s^2}\varrho_t \varrho_y\Big) + \\ \qquad + v_y^2\Big(\varrho_{tt} - 2\dfrac{\varrho_{st}}{\varrho_s}\varrho_t + \dfrac{\varrho_{ss}}{\varrho_s^2}\varrho_t^2\Big) + \\ \qquad + \varrho_{yy} - 2\dfrac{\varrho_{ys}}{\varrho_s}\varrho_y + \dfrac{\varrho_{ss}}{\varrho_s^2}\varrho_y^2\Big|^{13}), \end{cases} \tag{20}$$

wobei in den Ableitungen von ϱ die Argumente (x, y, s, v) stehen sollen[14]*. Über die in den Formeln* (15) *bis* (20) *stehenden Funktionen möge folgendes gelten: Die Funktionen $\varepsilon(x)$ und $\vartheta(x, y)$ seien definiert für $x \in \langle x_0, x_1 \rangle$ bzw. $(x, y) \in \Gamma_*$,*

[12]) Bei dieser heuristischen Überlegung werde davon abgesehen, daß hinter der Stelle $\xi + 1/\eta$ ein neuer Zweig der Funktion $z(x)$ beginnt, der mit dem Anfangswert $z(\xi) = \eta$ zunächst nichts zu tun hat. Immerhin sieht man bei der Transformation der DGl (14) auf die neue abhängige Veränderliche $1/z$, daß die angegebene Lösung die „natürliche" Fortsetzung über diese Singularität hinweg ist.

[13]) In der $m \times m$-Matrix $B(s)$ sind die $m \times m$-Matrizen $v_y \varrho_{yt}, \ldots, v_y v_y, \ldots$ als dyadische Produkte der Vektoren v_y und $\varrho_{yt}, \ldots, v_y$ und $v_y, \ldots$ aufzufassen.

[14]) Ist die Funktion $r(x, y)$ auf B mit Hilfe einer Funktion $c(x)$ implizit definiert durch $\varrho(x, y, r, v) = c(x)$, so ist der Vektor $A(r) = r_y$, die Matrix $B(r) = r_{yy}$.

es sei $\delta(x) \in C_0$ *in* $\langle x_0, x_1\rangle$ *und* $\omega(x, r) \in C_0$ *für* $x \in \langle x_0, x_1\rangle$, r *beliebig. Wird in* $\langle x_0, x_1\rangle$ *das Maximalintegral der gewöhnlichen expliziten DGl erster Ordnung*

$$z' = \omega(x, z) + \delta(x) \quad \textit{mit} \quad z(x_0) = \varepsilon(x_0) \tag{8}$$

mit $\sigma(x)$ *bezeichnet*[9]*, so sei*[10])

$$\sigma(x) \geqq \varepsilon(x) \quad \textit{in} \quad \langle x_0, x_1\rangle \tag{9}$$

und

$$\begin{cases} \operatorname{sign}(s - v)\,[\varrho_y(x, y, s, v)\, y_* + \varrho_s(x, y, s, v)\, h(x, y, s) + \\ \qquad + \varrho_t(x, y, s, v)\, k(x, y, v)] > \vartheta(x, y) \end{cases} \tag{21}$$

für $(x, y) \in \Gamma_*$ *und alle* $s \in \langle \alpha(x, y), \beta(x, y)\rangle$ *mit der Eigenschaft* $|\varrho(x, y, s, v)| > \sigma(x)$. *Schließlich gelte* $\varrho(x, y, s, t) \in C_0$ *für* $\alpha(x, y) \leqq s \leqq \beta(x, y)$, $\operatorname{Min} v(x, y) \leqq t \leqq \operatorname{Max} v(x, y)$, $(x, y) \in B$ *und dort sei* $\varrho(x, y, t, t) = 0$; *in* $\alpha < s < \beta$, $\operatorname{Min} v \leqq t \leqq \operatorname{Max} v$, $(x, y) \in G + \Gamma_1$ *sei* $\varrho \in C_2$[15]) *und dort gelte noch* $\varrho_s > 0$, $\varrho_t < 0$.

Die Schranken $\alpha(x, y)$ *und* $\beta(x, y)$ *seien dabei so weit, daß*

$$\varrho(x, y, \alpha, v) < -\sigma(x) \tag{22a}$$

und

$$\varrho(x, y, \beta, v) > \sigma(x) \tag{22b}$$

gilt für $(x, y) \in G + \Gamma_1$.

Unter diesen Voraussetzungen ist

$$|\varrho(x, y, z, v)| \leqq \sigma(x) \tag{23}$$

auf ganz B *für jede Lösung* $z \in Z$ *des Anfangs-Randwertproblems* AR.

Beweis. Zu vorgegebenem $\varepsilon_1 > 0$ werde das Maximalintegral der DGl

$$z'(x) = \omega(x, z) + \delta(x) + \varepsilon_1 \quad \text{mit} \quad z(x_0) = \varepsilon(x_0) + \varepsilon_1$$

gebildet und mit $\sigma_1(x)$ bezeichnet. Es sei ε_1 so klein, daß noch $\varrho(x, y, \beta, v) > \sigma_1$ und $\varrho(x, y, \alpha, v) < -\sigma_1$ bleibt in $G + \Gamma_1$.

Es sei $z \in Z$ eine Lösung des Anfangs-Randwertproblems AR. Angenommen, es gebe Stellen $(x', y') \in G + \Gamma_1$ mit $|\varrho(x', y', z, v)| > \sigma_1(x)$. Wegen der Beschränktheit von G gibt es daraus eine konvergente Folge (x'_n, y'_n) mit $x'_n \to \bar{x} = \inf x'$, $y'_n \to \bar{y}$. Wegen der Stetigkeit von z, v, ϱ und σ_1 und wegen (9) gilt dann $|\varrho(\bar{x}, \bar{y}, \bar{z}, \bar{v})| \geqq \sigma_1(\bar{x}) > \sigma(\bar{x}) \geqq \varepsilon(\bar{x})$, wenn $\bar{z} = z(\bar{x}, \bar{y})$, $\bar{v} = v(\bar{x}, \bar{y})$ ist. Wegen (15) liegt daher $(\bar{x}, \bar{y})$ nicht auf Γ^*, d.h. $(\bar{x}, \bar{y}) \in G + \Gamma_1 + \Gamma_*$. Damit ist eine Annäherung von kleineren x-Werten her möglich mit dem Resultat $|\varrho| \leqq \sigma_1(\bar{x})$, also gilt $|\varrho(\bar{x}, \bar{y}, \bar{z}, \bar{v})| = \sigma_1(\bar{x})$. Es werde zunächst

$$\varrho(\bar{x}, \bar{y}, \bar{z}, \bar{v}) = \sigma_1(\bar{x}) \tag{24}$$

[15]) Von dem trivialen Fall $v = \text{const}$, in dem z.B. die partielle Ableitung ϱ_t nicht erklärt zu sein braucht, werde abgesehen.

angenommen. Dann hat $\varrho(x, y, z, v) - \sigma_1(x)$ an der Stelle $(\bar{x}, \bar{y})$ ein relatives Maximum in $(x, y) \in G + \Gamma_1 + \Gamma_\nu$, $x \leqq \bar{x}$, d.h. in $(\bar{x}, \bar{y})$ ist

$$\frac{d}{d\nu}(\varrho - \sigma_1) \leqq 0, \quad \text{falls} \quad (\bar{x}, \bar{y}) \in \Gamma_\nu \tag{25}$$

gilt. Ist jedoch $(\bar{x}, \bar{y}) \in G + \Gamma_1$, so gilt dort

$$\frac{d}{dx}(\varrho - \sigma_1) \geqq 0, \tag{26}$$

$$\frac{d}{dy}(\varrho - \sigma_1) = 0 \qquad \frac{d^2}{dy^2}(\varrho - \sigma_1) \leqq 0 \text{ [16]}. \tag{27}$$

Angenommen, es sei $(\bar{x}, \bar{y}) \in \Gamma_\nu$. Nach (25), (16) und (21) wäre dann dort

$$0 \geqq \frac{d}{d\nu}(\varrho - \sigma_1) = \varrho_y y_\nu + \varrho_z z_\nu + \varrho_t v_\nu \geqq \varrho_y y_\nu + \\ + \varrho_z h(x, y, z) + \varrho_t k(x, y, v) - \vartheta > 0 \text{ [17]},$$

also kann $(\bar{x}, \bar{y})$ nicht auf Γ_ν liegen, d.h. $(\bar{x}, \bar{y}) \in G + \Gamma_1$.

Nach (26) ist dann in $(\bar{x}, \bar{y})$:

$$0 \leqq \frac{d}{dx}(\varrho - \sigma_1) = \varrho_x + \varrho_z z_x + \varrho_t v_x - \sigma_1'$$

und wegen (2), (17), (8) und (24) gilt

$$\begin{cases} 0 \leqq \varrho_x + \varrho_z f(x, y, z, z_y, z_{yy}) + \varrho_t g(x, y, v, v_y, v_{yy}) + \delta(x) - \omega(x, \sigma_1) - \\ \qquad - \delta(x) - \varepsilon_1 \\ < \varrho_x + \varrho_z f(x, y, z, z_y, z_{yy}) + \varrho_t g(x, y, v, v_y, v_{yy}) - \omega\big(x, \varrho(x, y, z, v)\big). \end{cases} \tag{28}$$

Nun ist nach (24) aber $\bar{z} > \bar{v}$ und wegen (27), (19) und (20) gilt bei $(\bar{x}, \bar{y})$ die Beziehung $z_y = A(z)$, $z_{yy} \leqq B(z)$. In (18) eingesetzt gibt dies wegen des monotonen Wachstums von f bezüglich der letzten Veränderlichen

$$\varrho_x + \varrho_z f(x, y, z, z_y, z_{yy}) + \varrho_t g(x, y, v, v_y, v_{yy}) - \omega\big(x, \varrho(x, y, z, v)\big) \leqq 0$$

im Gegensatz zu (28). Aus diesem Widerspruch folgt $\varrho(x, y, z, v) \leqq \sigma_1(x)$ auf ganz B und wegen $\sigma_1(x) \to \sigma(x)$ für $\varepsilon_1 \to 0$ gilt auch noch $\varrho(x, y, z, v) \leqq \sigma(x)$ auf B. Genau so beweist man auch $\varrho(x, y, z, v) \geqq -\sigma(x)$, womit die behauptete Abschätzung (23) vollständig bewiesen ist.

Bemerkungen. Einzigkeit: Besitzt für $f \equiv g$, $h \equiv k$ und unter den Voraussetzungen des Satzes 2 die DGl $z' = \omega(x, z)$ unter der Anfangsbedingung $z(x_0) = 0$ nur die Lösung $\sigma_0(x) \equiv 0$, so hat das Anfangs-Randwertproblem AR offenbar höchstens *eine* Lösung.

Stabilität: Ebenfalls läßt sich leicht aus Aussagen über das Wachstum des Maximalintegrals $\sigma(x)$ der DGl (8) auf die Stabilität von Lösungen des Anfangs-Randwertproblems AR für $x \to \infty$ schließen, wenn das Gebiet G nach

[16] Das heißt: die Matrix $\frac{d^2}{dy^2}(\varrho - \sigma_1)$ ist negativ semidefinit.

[17] Der Ausdruck $\varrho_y y_\nu$ ist als skalares Produkt aufzufassen.

rechts nicht beschränkt ist und wenn die Abstandsfunktion $\varrho(x, y, s, t)$ für große Werte von x gewissen Wachstumsbedingungen genügt.

Verbesserung der Abschätzung: Um zu einer günstigen Abschätzung (23) zu gelangen, um also eine möglichst kleine Schranke $\sigma(x)$ zu erhalten, ist nach (8) neben einem möglichst kleinen Betrag des Defekts $\delta(x)$ und des Randfehlers $\varepsilon(x)$ eine möglichst kleine Funktion $\omega(x, r)$ anzustreben, und zwar sollte möglichst $\omega(x, r) \leqq 0$ sein. Daß man dies häufig durch passende Wahl der Abstandsfunktion $\varrho(x, y, s, t)$ erreichen kann, soll an einigen Beispielen gezeigt werden (vgl. insbesondere Beispiel 2 und 4). Aber auch bei festgewähltem ϱ hängt die Funktion $\omega(x, r)$ im allgemeinen noch wesentlich von der Näherungslösung $v(x, y)$ ab. Unter Umständen ist es daher vorteilhafter, bei der Wahl von $v(x, y)$ etwas ungünstigere Werte des Defekts $\delta(x)$ und des Randfehlers $\varepsilon(x)$ zuzulassen, um durch eine kleinere Funktion $\omega(x, r)$ trotzdem zu einer besseren Schranke $\sigma(x)$ zu kommen (man vgl. hierzu Beispiel 4). Schließlich sind die folgenden Beispiele noch so ausgewählt, daß bei ihnen nicht alle Voraussetzungen des Satzes 2 erfüllt sind. Obzwar nämlich die Aussagen dieses Satzes recht allgemein gefaßt sind, ist es doch in Sonderfällen möglich, unter noch schwächeren Forderungen die Gültigkeit der Abschätzung (23) sicherzustellen. Wollte man diese Tatsache in die Formulierung des Satzes 2 übernehmen, so würde er dadurch erheblich unhandlicher werden. Es ist daher zweckmäßiger, die Auswirkungen von Abschwächungen der Annahmen auf das Ergebnis (23) von Fall zu Fall zu untersuchen.

Beispiel 2. Gegeben sei für $m = 1$ die Wärmeleitungsgleichung

$$Tz \equiv z_x - z_{yy} = 0 \qquad \text{in } G + \Gamma_1$$

mit den Randwerten

$$R^* z \equiv z - \zeta = 0 \qquad \text{auf } \Gamma^*,$$

Γ_1 sei leer.

a) Für $f \equiv g$,

$$\varrho(x, y, s, t) = s - t$$

ergibt sich $\omega(x, r) = 0$ nach Satz 1, und daher die Abschätzung

$$|z - v| \leqq \sigma(x) = \varepsilon(x_0) + \int_{x_0}^{x} \delta(t)\, dt \tag{29}$$

für $\sigma(\dot{x}) \geqq \varepsilon(x)$.

b) Betrachtet man dagegen (wieder für $f \equiv g$) den Abstand

$$\varrho(x, y, s, t) = \frac{s - t}{w(x, y)}$$

mit $w > 0$, $w \in C_2$ auf B, so geht (15) über in

$$|R^* v| \leqq w(x, y)\, \varepsilon(x),$$

aus (17) wird

$$|Sv| \leqq w(x, y)\, \delta(x)$$

und aus (18) folgt die Bedingung

$$\frac{1}{w}(-w_x + w_{yy})\frac{|s-v|}{w} \leqq \omega\left(x, \frac{|s-v|}{w}\right).$$

Diese ist erfüllt für

$$\omega(x, r) = l(x)\, r \tag{30}$$

mit $l \in C_0$ und

$$\frac{-Tw}{w} \equiv \frac{1}{w}[-w_x + w_{yy}] \leqq l(x).$$

Die DGl (8) ist damit eine lineare DGl mit der Lösung

$$\sigma(x) = \exp\int_{x_0}^{x} l(t)\,dt\left[\varepsilon(x_0) + \int_{x_0}^{x}\delta(x)\exp -\int_{x_0}^{t} l(\tau)\,d\tau\,dt\right] \tag{31}$$

und mit dieser Funktion geht (23) über in

$$|z - v| \leqq w(x, y)\,\sigma(x). \tag{32}$$

Es leuchtet ein, daß durch passende Funktionen $w(x, y)$ die Abschätzung (32) gegenüber (29) verbessert werden kann.

Zahlenbeispiel. Es sei $G: \{x_0 < x < x_1,\ -1 < y < 1\}$, auf

$$\Gamma^*: \begin{cases} x = x_0, |y| \leqq 1 \\ x_0 < x \leqq x_1, |y| = 1 \end{cases}$$

seien die Randwerte

$$\zeta(x, y) = \begin{cases} 1 - y^2 & \text{für } x = x_0, |y| \leqq 1 \\ 0 & \text{für } x_0 < x \leqq x_1, |y| = 1 \end{cases}$$

vorgeschrieben. Weiter gelte $f \equiv g$, $v(x, y) = 1{,}04\, e^{-\frac{\pi^2}{4}(x - x_0)} \cos\frac{\pi}{2} y$, $w(x, y) = 1 - y^2$ und damit $\frac{-Tw}{w} = \frac{-2}{1-y^2} \leqq -2 \equiv l(x)$.

Die Ergebnisse unter diesen Annahmen sind in Tabelle 1 aufgezeichnet. Obwohl im Falle b) der maximale Randfehler $\varepsilon(x)$ fast fünfmal so groß ist wie im Falle a), ist die Abschätzung (23) durch die günstigere Funktion $\omega(x, r)$ — wenn man von dem Gebiet $0 \leqq x - x_0 < 0{,}78$ absieht — erheblich besser im Falle b) als bei a). Insbesondere sieht man daraus, daß der Betrag $|z - v|$ des Fehlers für $x \to \infty$ und für $|y| \to 1$ verschwindet, was aus der Abschätzung bei a) nicht zu entnehmen ist. Schließlich muß noch festgestellt werden, daß im Falle b) der Satz 2 nicht ohne weiteres anwendbar ist, da ja $\varrho(x, y, s, t)$ auf Γ^* nicht mehr stetig ist. Verwendet man jedoch die Funktion $\overline{w}(x, y) = \bar{\varepsilon} + 1 - y^2$ mit $\bar{\varepsilon} > 0$, dann sind alle Voraussetzungen des Satzes 2 erfüllt und es ergibt sich

$$|z - v| \leqq 0{,}19\,(\bar{\varepsilon} + 1 - y^2)\, e^{-2(x - x_0)}.$$

Für $\bar{\varepsilon} \to 0$ ergibt sich daraus die behauptete Abschätzung von Tabelle 1.

Beispiel 3. Es sei $m = 1$,

$$(33)\quad f(x, y, r, s, t) \equiv g(x, y, r, s, t) = a(x, y)\, r + b(x, y)\, s + c^2(x, y)\, |r|^{\varkappa} t$$

mit $-\infty < \varkappa < +\infty$; a, b, c definiert in $G + \Gamma_1$. Der triviale Fall $\varkappa = 0$ der linearen DGl sei im folgenden ausgeschlossen. Man betrachtet also die Klasse der parabolischen DGlen

$$Tz \equiv z_x - a(x, y)\, z - b(x, y)\, z_y - c^2(x, y)\, |z|^{\varkappa} z_{yy} = 0 \qquad \text{in } G + \Gamma_1$$

Tabelle 1

Formel-Nr.	Allgemeine Gestalt	Fall a) $\varrho(x, y, s, t) = s - t$	Fall b) $\varrho(x, y, s, t) = \frac{s-t}{1-y^2}$
(15)	$\lvert\varrho(x, y, \zeta, v)\rvert \leqq \varepsilon(x)$	$\lvert\zeta - v\rvert \leqq \varepsilon(x)$ $= \begin{cases} 0{,}04 & \text{für } x = x_0 \\ 0 & \text{für } x > x_0 \end{cases}$	$\frac{\lvert\zeta - v\rvert}{1 - y^2} \leqq \varepsilon(x)$ $= \begin{cases} 0{,}19 & \text{für } x = x_0 \\ 0 & \text{für } x > x_0 \end{cases}$
(17)	$\lvert\varrho_t(x, y, s, v)\ Sv\rvert \leqq \delta(x)$	$Sv \equiv 0,\ \delta(x) \equiv 0$	
(18)	$\cdots \leqq \omega(x, \lvert\varrho(x, y, s, v)\rvert)$ liefert die Funktion $\omega(x, r)$	$0 \leqq \omega(x, \lvert s - t\rvert)$. $\omega(x, r) = 0$	$\frac{-2}{1-y^2}\lvert\varrho\rvert \leqq \omega(x, \lvert\varrho\rvert)$ $\omega(x, r) = -2r$
(8)	$z' = \omega(x, r) + \delta(x)$, $z(x_0) = \varepsilon(x_0)$ besitzt das Maximalintegral $\sigma(x)$	$z' = 0,\ z(x_0) = 0{,}04$ $\sigma(x) = 0{,}04$	$z' = -2z,\ z(x_0) = 0{,}19$ $\sigma(x) = 0{,}19\, e^{-2(x - x_0)}$
(23)	$\lvert\varrho(x, y, z, v)\rvert \leqq \sigma(x)$	$\lvert z - v\rvert \leqq 0{,}04$	$\lvert z - v\rvert \leqq$ $\leqq 0{,}19\,(1 - y^2)\, e^{-2(x - x_0)}$

unter Randbedingungen $Rz = 0$ auf Γ. Es ist zweckmäßig, hierbei die Abstandsfunktion

$$\varrho(x, y, s, t) = \operatorname{sign}\varkappa\, [\lvert s\rvert^{\varkappa} - \lvert t\rvert^{\varkappa}]$$

zu benutzen. Für $\alpha(x, y) \equiv 0 \leqq \lvert s\rvert, \lvert t\rvert < \infty \equiv \beta(x, y)$ ist $\varrho \in C_0$ und $\varrho(x, y, t, t) \equiv 0$, in $0 < \lvert s\rvert, \lvert t\rvert < \infty$ ist sogar $\varrho \in C_\infty$. Weiter gilt dort noch $\varrho_s = \lvert\varkappa\rvert\, \lvert s\rvert^{\varkappa - 1} > 0$, $\varrho_t = -\lvert\varkappa\rvert\, \lvert t\rvert^{\varkappa - 1} < 0$. Der Einfachheit halber sei im folgenden $z \geqq 0$ und $v \geqq 0$ vorausgesetzt, damit können in ϱ und f die Betragstriche entfallen. Diese Voraussetzung ist für jede Lösung $z \in Z$ von AR nach dem Maximum-Minimumprinzip z. B. erfüllt, wenn $\Gamma = \Gamma^*$ ist und nur Randwerte $\zeta \geqq 0$ auf Γ^* vorgeschrieben sind (vgl. K. Nickel [9], S. 12/13). Setzt man nun f und ϱ in (18) ein, so ergibt sich die einfache Ungleichung

$$\left[\lvert\varkappa\rvert\, a(x, y) + \operatorname{sign}\varkappa\, c^2(x, y)\, [v^{\varkappa}]_{yy}\right] \lvert s^{\varkappa} - v^{\varkappa}\rvert \leqq \omega\left(x, \lvert s^{\varkappa} - v^{\varkappa}\rvert\right)$$

für $s, v \geqq 0$. Setzt man wie in Beispiel 2b)

$$(30)\qquad \omega(x, r) = l(x)\, r$$

mit $l(x) \in C_0$, so ist (18) erfüllt, falls nur

$$|\varkappa|\, a(x,y) + \operatorname{sign}\varkappa\, c^2(x,y)\,[v^\varkappa]_{yy} \leqq l(x) \tag{34}$$

bleibt in $G+\Gamma_1$. Eine solche Funktion $l(x)$ läßt sich stets finden, wenn etwa $a, c \in C_0$ und $[v^\varkappa]_{yy} \in C_0$ vorausgesetzt wird auf ganz B. Mit (30) wird (8) wieder eine lineare DGl mit der Lösung $\sigma(x)$ nach (31) und damit lautet die Abschätzung (23) in diesem Falle

$$|z^\varkappa - v^\varkappa| \leqq \sigma(x). \tag{35}$$

Wegen $\sigma(x) \to 0$ für $\varepsilon(x) \to 0$ und $\delta(x) \to 0$ ergibt sich daraus insbesondere die Einzigkeit einer Lösung $z \in Z$ des betrachteten Anfangs-Randwertproblems, falls noch die restlichen Bedingungen von Satz 2 erfüllt sind.

Zahlenbeispiel 4. In (33) sei $x_0 = 0$, $a(x,y) = b(x,y) = 0$, $c(x,y) = 1$, $\varkappa = \frac{1}{2}$,

$G: \{0 = x_0 < x < x_1,\ -1 < y < 1\}$, Γ_1 sei leer, auf Γ^*: $\begin{cases} x = 0,\ |y| \leqq 1 \\ 0 < x \leqq x_1,\ |y| = 1 \end{cases}$

seien die Randwerte

$$\zeta(x,y) = \begin{cases} \cos^4 \dfrac{\pi}{2} y & \text{für} \quad x = 0,\ |y| \leqq 1 \\ 0 & \text{für} \quad 0 < x \leqq x_1,\ |y| = 1 \end{cases}$$

vorgeschrieben. Man betrachtet also das Anfangs-Randwertproblem

$$\text{AR:} \begin{cases} Tz \equiv z_x - \sqrt{z}\, z_{yy} = 0 & \text{in } G+\Gamma_1, \\ R^* z \equiv z - \zeta = 0 & \text{auf } \Gamma. \end{cases} \tag{36}$$

a) Man setzt $v(x,y) = 0{,}9025\, \dfrac{(1-y^2)^4}{(x+1)^2}$ und erhält aus (15) die Abschätzung

$$|\sqrt{\zeta} - \sqrt{v}| \leqq \varepsilon(x) = \begin{cases} 0{,}05 & \text{für} \quad x = 0,\ |y| \leqq 1 \\ 0 & \text{für} \quad x > 0,\ |y| = 1. \end{cases}$$

Einsetzen von v in (17) liefert

$$|\varrho_t S v| \equiv \frac{1}{2\sqrt{v}}\, |v_x - \sqrt{v}\, v_{yy}| \leqq \frac{2{,}88}{(x+1)^2} = \delta(x).$$

Die Funktion $l(x)$ berechnet sich nach (34) aus

$$[\sqrt{v}]_{yy} = -\frac{3{,}8}{x+1}\,(3y^2 - 1) \leqq \frac{7{,}6}{x+1} \equiv l(x).$$

Mit den Funktionen $\varepsilon(x)$, $\delta(x)$ und $l(x)$ ergibt sich aus (31) und nach (35) die Abschätzung

$$\left| \sqrt{z} - \frac{0{,}95}{x+1}\,(1-y^2)^2 \right| \leqq \sigma(x) \leqq 0{,}385\,(x+1)^{7{,}6} - \frac{0{,}335}{x+1} \tag{37}$$

für jede Lösung $z \in Z^*$ von (36). Obwohl bei der obenstehenden Näherungsfunktion v der Randfehler $\varepsilon(x)$ sehr klein wird und der Defekt $\delta(x)$ für $x \to \infty$ gegen Null strebt, ist die Abschätzung (37) für große Werte von x unbrauchbar

(vgl. Fig. 2). Die Ursache dafür liegt in der ungünstigen Funktion $\omega(x, r)$. Bevor jedoch ω und damit die Abschätzung (37) durch eine günstigere Wahl von $v(x, y)$ verbessert wird, sollen noch die Bedingungen (22) nachgeprüft werden. Es ist $\alpha(x, y) = 0$, $\beta(x, y) = \infty$, also ist zwar (22b) erfüllt, aber (22a) ist wegen $\varrho(x, y, \alpha, v) = 0 - \sqrt{v} = -\frac{0{,}95}{x+1}(1-y^2)^2 \not\leqq -\sigma(x)$ in $G + \Gamma_1$ verletzt. Tatsächlich ist jedoch die Aussage des Satzes 2 und damit (37) trotzdem noch richtig, wenn nur außer $z \in Z$ noch $\sqrt{z} \in Z$ vorausgesetzt wird (dies ist eine zusätzliche Bedingung allein an den Stellen $(x, y) \in B$ mit $z(x, y) = 0$). Beim

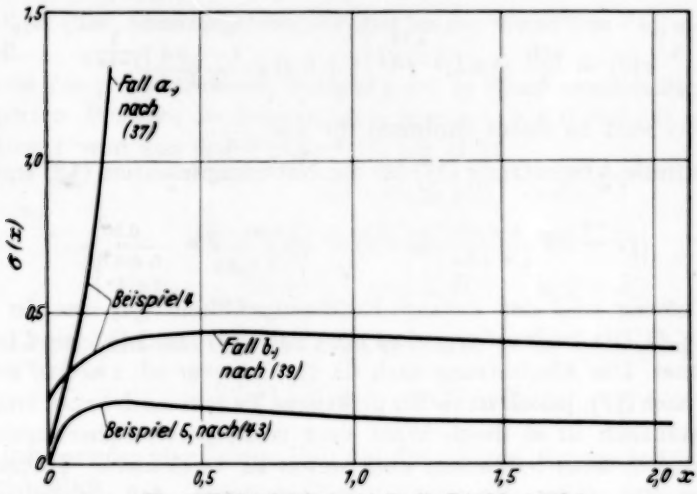


Fig. 2. Fehlerschranken $\sigma(x)$ für die Beispiele 4 und 5

Beweis des Satzes 2 stellt nämlich die Bedingung (22a) im Falle $\varrho(\bar{x}, \bar{y}, \bar{z}, \bar{v}) = -\sigma_1(\bar{x})$ sicher, daß die Funktion $\varrho(x, y, z, v) = \sqrt{z} - \sqrt{v}$ an der Stelle $(\bar{x}, \bar{y}) \in G + \Gamma_1$ partiell differenziert werden darf. Für $\sqrt{z} \in Z$ ist das erfüllt und der Beweis läßt sich wie angegeben zu Ende führen.

b) Zur Verbesserung der Abschätzung (37) setzt man

$$v(x, y) = 0{,}64\,\lambda^2 \frac{(1-y^2)^2}{(x+\lambda)^2} \tag{38}$$

mit einem vorläufig noch freien Parameter $\lambda > 0$. Aus (15), (17) und (34) ergeben sich der Reihe nach

$$\varepsilon(x) = \begin{cases} 0{,}21 & \text{für} \quad x = 0 \\ 0 & \text{für} \quad x > 0 \end{cases}$$

$$\delta(x) = \frac{\varphi(\lambda)}{(x+\lambda)^2}$$

mit

$$\varphi(\lambda) = \begin{cases} 0{,}8\,\lambda(1 - 1{,}6\,\lambda) & \text{für} \quad 0 < \lambda \leqq \frac{1}{4{,}8} \\ 2{,}56\,\lambda^2 & \text{für} \quad \lambda \geqq \frac{1}{4{,}8} \end{cases}$$

und

$$l(x) = -\frac{1{,}6\lambda}{x+\lambda}.$$

Die Auflösung der DGl (8) ist damit nach (31):

$$\sigma(x) = \left[0{,}21 + \frac{\varphi(\lambda)}{\lambda(1-1{,}6\lambda)}\right]\left(\frac{\lambda}{x+\lambda}\right)^{1{,}6\lambda} - \frac{\varphi(\lambda)}{1-1{,}6\lambda}\frac{\lambda}{x+\lambda}.$$

$\sigma(x)$ nimmt seinen Maximalwert für $x \geqq 0$ an der Stelle

$$\bar{x} = -\lambda + \lambda(1{,}6\lambda)^{\frac{1}{1{,}6\lambda-1}}\left[1 + 0{,}21\lambda\frac{1-1{,}6\lambda}{\varphi(\lambda)}\right]^{\frac{1}{1{,}6\lambda-1}}$$

an, es ist

$$\sigma(\bar{x}) = \frac{\varphi(\lambda)}{\lambda}(1{,}6\lambda)^{\frac{1{,}6\lambda}{1-1{,}6\lambda}}\left[1 + 0{,}21\lambda\frac{1-1{,}6\lambda}{\varphi(\lambda)}\right]^{\frac{1}{1-1{,}6\lambda}}.$$

Dieser Wert wird zu einem Minimum für $\lambda = \frac{1}{4{,}8}$.

Als optimale Abschätzung (23) für die Näherungsfunktion (38) ergibt sich daher

$$\left|\sqrt{z} - 0{,}8\frac{1-y^2}{1+4{,}8x}\right| \leqq \sigma(x) = \frac{1{,}01}{\sqrt[3]{1+4{,}8x}} - \frac{0{,}80}{1+4{,}8x} \tag{39}$$

für jede Lösung $z \in Z$ des Anfangs-Randwertproblems (36) mit der Eigenschaft $\sqrt{z} \in Z$. Die beiden Schranken $\sigma(x)$ nach (37) und (39) sind in Fig. 2 aufgezeichnet. Die Abschätzung nach Gl. (39) ist zwar ab $x \approx 0{,}07$ erheblich besser als nach (37), jedoch ist sie für praktische Zwecke noch kaum brauchbar. Selbstverständlich ist es durch Wahl einer verbesserten Näherungsfunktion $v(x, y)$ möglich, diese Schranken noch weiter zu verkleinern. Immerhin ist jedoch aus (39) wegen $\lim_{x\to\infty}\sigma(x) = 0$ zu entnehmen, daß alle betrachteten Lösungen von (36) für $x \to \infty$ verschwinden.

Es zeigt sich aus diesem Beispiel, daß es sich lohnen kann, bei der Wahl der Näherungsfunktion $v(x, y)$ einen etwas höheren Randfehler $\varepsilon(x)$ (und/oder einen größeren Defekt $\delta(x)$) in Kauf zu nehmen, um dafür durch eine günstigere Funktion $\omega(x, r)$ trotzdem zu einer besseren Abschätzung (23) zu gelangen.

Schließlich sei noch bemerkt, daß man die Abschätzungen (37) und (39) statt in der Gestalt $|\sqrt{z} - \sqrt{v}| \leqq \sigma(x)$ auch in der Gestalt $|z - v - \sigma^2(x)| \leqq 2\sqrt{v}\,\sigma(x)$ schreiben kann.

Beispiel 5. (Croccosche Grenzschicht-DGl). Nach L. CROCCO [11] (vgl. auch K. NICKEL [9], S. 19/20) läßt sich die laminare stationäre und zweidimensionale Wandgrenzschicht in einem inkompressiblen Medium durch das folgende Anfangs-Randwertproblem beschreiben: Es sei $m = 1$,

$$G: \{x_0 < x < x_1,\ 0 < y < U(x)\},$$

$$\Gamma^*: \begin{Bmatrix} x = x_0,\ 0 \leqq y \leqq U(x_0) \\ x_0 < x \leqq x_0,\ y = U(x) \end{Bmatrix}, \qquad \Gamma_*: \{x_0 < x \leqq x_1,\ y = 0\},$$

$v \equiv y$, die „Außengeschwindigkeit" $U(x) > 0$ sei stetig und differenzierbar in $\langle x_0, x_1\rangle$. Die Funktion f besitzt die Gestalt (33) mit $a \equiv 0$, $b = \frac{UU'}{y}$, $c^2 = \frac{1}{y}$,

$\varkappa = 2$, damit lautet die Croccosche DGl

$$Tz \equiv z_x - \frac{x^2 z_{yy}}{y} + \frac{UU'}{y} z_y = 0 \quad \text{in } G + \Gamma_1.$$

Die Randbedingungen sind

$$R^* z \equiv z - \zeta = 0 \text{ auf } \Gamma^* \text{ mit } \zeta(x, y) = \begin{cases} \bar{z}(y) & \text{für } x = x_0,\ 0 \leqq y \leqq U(x_0) \\ 0 & \text{für } x_0 < x \leqq x_1,\ y = U(x) \end{cases}$$

und

$$R_\nu z \equiv z_\nu - h(x, y, z) = 0 \text{ auf } \Gamma_\nu \text{ mit } h(x, y, z) = V_0(x) - \frac{U(x)\, U'(x)}{z},$$

wobei $V_0(x)$ (die „Absaugegeschwindigkeit an der Wand") in $\langle x_0, x_1 \rangle$ definiert sein soll.

Es sei $f \equiv g$, $h \equiv k$. Nach Beispiel 3 ist es dann zweckmäßig, bei dem vorliegenden Problem die Abstandsfunktion $\varrho(x, y, s, t) = s^2 - t^2$ zu verwenden. Damit wird aus den UnGlen (15) bis (17):

$$\text{(40)} \qquad \begin{cases} |\zeta^2 - v^2| \leqq \varepsilon(x) & \text{auf } \Gamma^*, \\ 2|v(v_y - V_0) + UU'| \leqq \vartheta(x) & \text{auf } \Gamma_\nu, \\ 2\left|v v_x - \frac{v^3 v_{yy}}{y} + \frac{UU'}{y} v v_y\right| \leqq \delta(x) & \text{in } G + \Gamma_1. \end{cases}$$

Die UnGl (18) geht mit (30) und nach (34) über in die Bedingung

$$\text{(41)} \qquad \frac{1}{y}[v^2]_{yy} \leqq l(x) \quad \text{in } G + \Gamma_1,$$

(8) ist dann wieder eine lineare DGl mit der Lösung $\sigma(x)$ nach (31). Aus (21) wird schließlich

$$\text{(42)} \qquad 2V_0|s - v| > \vartheta(x) \quad \text{auf } \Gamma_\nu$$

für $|s^2 - v^2| > \sigma(x) \geqq 0$. Ist $\vartheta(x) \equiv 0$, dann ist (42) sicher erfüllt für $V_0(x) > 0$ („Ausblasen" an der Wand). Obzwar im Grenzfall $V_0(x) \equiv 0$ („undurchlässige Wand") die Bedingung (42) verletzt ist, bleibt die Aussage (23) des Satzes 2 bei dem hier vorliegenden sehr „einfachen" Randstück Γ_ν trotzdem noch gültig, falls man nur voraussetzt, daß (17) und (18), d. h. (40) und (41), statt auf $G + \Gamma_1$ sogar auf $G + \Gamma_1 + \Gamma_\nu$ erfüllt sind und wenn man außerdem nur solche Lösungen $z \in Z$ des Anfangs-Randwertproblems AR zuläßt, die der Gleichung $Tz = 0$ sogar auf $G + \Gamma_1 + \Gamma_\nu$ genügen. Beim Beweis von Satz 2 läßt sich dann zwar die Negation $(\bar{x}, \bar{y}) \notin \Gamma_\nu$ nicht mehr beweisen, doch ergibt sich offenbar wegen $\frac{\partial}{\partial \nu} \equiv \frac{\partial}{\partial y}$ unter den angegebenen zusätzlichen Voraussetzungen der gewünschte Widerspruch aus (28) auch schon mit $(\bar{x}, \bar{y}) \in G + \Gamma_1 + \Gamma_\nu$. Im folgenden soll dieser Grenzfall $V_0(x) \equiv 0$ in dem Sonderfall $U'(x) \equiv 0$ an einem Zahlenbeispiel gezeigt werden.

Zahlenbeispiel. Es sei $x_0 = 0$, $U(x) \equiv 1$, auf Γ^* seien die Anfangs-Randwerte

$$\zeta(x, y) = \begin{cases} \sqrt{1 - y^3} & \text{für } x = 0,\ 0 \leqq y \leqq 1 \\ 0 & \text{für } 0 < x \leqq x_1,\ y = 1 \end{cases}$$

vorgeschrieben. Man setzt

$$v(x, y) = \sqrt{\lambda \frac{1-y^3}{x+\lambda}} \quad \text{auf } B$$

mit einem vorläufig noch freien Parameter $\lambda > 0$ und erhält mit $\varepsilon(x) \equiv 0$, $\vartheta(x) \equiv 0$,

$$\delta(x) = \begin{cases} \dfrac{9\lambda^3}{2(x+\lambda)^3} & \text{für } \lambda \geqq \dfrac{2}{21} \\ \lambda \dfrac{(1-6\lambda)}{(x+\lambda)^2}. & \text{für } 0 \leqq \lambda \leqq \dfrac{2}{21}, \end{cases}$$

$l(x) = -\dfrac{6\lambda}{x+\lambda}$ auf gleiche Weise wie in Beispiel 4 den optimalen Wert $\lambda_{\text{opt}} = \dfrac{2}{21}$. Mit der Schrankenfunktion

$$\sigma(x) = \left(\frac{2}{2+21x}\right)^{\frac{1}{3}} - \frac{2}{2+21x}$$

(vgl. Fig. 2) ergibt sich daher die Abschätzung

$$(43) \qquad \left| z^2 - 2\frac{1-y^3}{2+21x} \right| \leqq \sigma(x) = \left(\frac{2}{2+21x}\right)^{\frac{1}{3}} - \frac{2}{2+21x} \quad \text{auf ganz } B$$

für jede Lösung $z \in Z$ von AR, für die $Tz = 0$ sogar noch auf Γ_2 gilt. Wegen $\lim_{x \to \infty} \sigma(x) = 0$ folgt daraus das Verschwinden jeder derartigen Lösung von AR, also die Stabilität. Dagegen ist die Abschätzung (43) zu grob, um zu zeigen, daß für $0 \leqq x < \infty$ stets $z(x, 0) > 0$ bleibt (d.h. daß „die Wandschubspannung nicht verschwindet und also die Strömung nicht ablöst"). Zum Beweis vgl. man K. NICKEL [9], S. 22/23.

Literatur

[1] NAGUMO, M.: Note in Kansū-Hōteisiki No. 15 (1939) [in japanisch]. Zitiert nach [3]. — [2] WESTPHAL, H.: Zur Abschätzung der Lösungen nichtlinearer parabolischer Differentialgleichungen. Math. Z. **51**, 690–695 (1949). — [3] NAGUMO, M., et S. SIMODA: Note sur l'inégalité différentielle concernant les équations du type parabolique. Proc. Japan Acad. **27**, 536–539 (1951). — [4] PRODI, G.: Questioni di stabilità per equazioni non lineari alle derivate parziali di tipo parabolico. Atti Accad. naz. Lincei, Rend., Cl. Sci. fis. mat. natur., VIII. Ser., **10**, 365–370 (1951). — [5] NARASIMHAN, R.: On the Asymptotic Stability of Solutions of Parabolic Differential Equations. J. Rat. Mech. Analysis **3**, 303–313 (1954). — [6] SZARSKI, J.: Sur la limitation et l'unicité des solutions d'un système non-linéaire d'équations paraboliques aux dérivées partielles du second ordre. Ann. Polon. Math. **2**, 237–249 (1955). — [7] COLLATZ, L.: Fehlerabschätzungen für Näherungslösungen parabolischer Differentialgleichungen. An. Acad. Brasil. Ci. **28**, No. 1 (1956). — [8] MLAK, W.: Differential inequalities of parabolic type. Ann. Polon. Math. **3**, 349–354 (1957). — [9] NICKEL, K.: Einige Eigenschaften von Lösungen der Prandtlschen Grenzschicht-Differentialgleichungen. Arch. Rat. Mech. Analysis **2**, 1–31 (1958). — [10] MÜLLER, M.: Über die Eindeutigkeit der Integrale eines Systems gewöhnlicher Differentialgleichungen und die Konvergenz einer Gattung von Verfahren zur Approximation dieser Integrale. S.-Ber. Heidelberger Akad. Wiss., math.-naturw. Kl. **1927**, 9. Abh. — [11] CROCCO, L.: Lo strato limite laminare nei gas. Monograf. sci. Aeronaut. No. 3, 1946. — [12] COLLATZ, L.: Fehlerabschätzungen bei Randwertaufgaben partieller Differentialgleichungen mit unendlichem Grundgebiet. Z. Angew. Math. Phys. **9a**, 118–128 (1958).

Institut f. Angew. Mathematik der Techn. Hochschule Karlsruhe

(Eingegangen am 5. Dezember 1958)

WEVER, F.
Math. Zeitschr. 71, 283–288 (1959)

Über stationäre Null-Eins-Prozesse mit beschränkten Lücken

Von
FRANZ WEVER

Es sei ein stationärer Prozess $\{x_n; n=0, 1, \ldots\}$ vorgegeben, dessen Zufallsvariablen x_n nur die Werte Eins und Null mit den stationären Wahrscheinlichkeiten p und $q=1-p$ annehmen können. Hierzu werde ein weiterer Prozeß $\{\lambda_t; t=1, 2, \ldots\}$ definiert: Wir setzen $x_0=1$ ohne Einschränkung der Allgemeinheit. Ist jetzt weiter $x_{n_t}=1$ für eine monoton zunehmende Folge von Indizes $0=n_0<n_1<\cdots<n_i<\cdots$ und $x_n=0$ sonst, dann heiße $\lambda_i=n_i-n_{i-1}$ die *Länge der i-ten Lücke* des Prozesses $\{x_n\}$. Die Lückenlängen λ_t sind selbst Zufallsvariable, deren Werte natürliche Zahlen sind.

Jeder Null-Eins-Prozeß $\{x_n\}$ bestimmt vermöge dieser Zuordnung einen Prozeß $\{\lambda_t\}$ und umgekehrt ist $\{x_n\}$ durch Vorgabe des Prozesses seiner Lücken eindeutig definiert.

Unabhängigkeit der x_n hat auch Unabhängigkeit der λ_t zur Folge, und zwar sind in diesem Fall die λ_t nach der geometrischen Verteilung verteilt. Andererseits bedeutet Unabhängigkeit der λ_t bei jeder von der geometrischen verschiedenen Verteilung stets eine Abhängigkeit der x_n. Hier sollen nun solche Null-Eins-Prozesse untersucht werden, deren Lücken voneinander unabhängig und beschränkt sind. Dazu werde $\{\lambda_t\}$ durch die Verteilung $p_i=p\{\lambda_t=i\}$, $i=1, \ldots, k, \sum_{i=1}^{k} p_i=1$ als Folge unabhängiger Zufallsvariabler definiert. Die Annahme einer maximalen Lückenlänge k steht praktischen Anwendungen meist nicht entgegen. Es werde stets $p_k>0$ vorausgesetzt; k soll also die Länge der größten Lücke sein, für die noch eine positive Wahrscheinlichkeit besteht. Die folgende Voraussetzung ist noch zweckmäßig und soll für die ganze Untersuchung gelten: Sind ν_j diejenigen Indizes, für die $p_{\nu_j}>0$, $j=1, \ldots, \varkappa, \nu_\varkappa=k$, dann soll der größte gemeinsame Teiler $GGT(\nu_1, \ldots, \nu_\varkappa)=1$ sein, d.h., der Prozeß $\{x_n\}$ soll aperiodisch sein. Das Auftreten von Lücken mit einem von Eins verschiedenen größten gemeinsamen Teiler ließe sich durch eine einfache Umformung des Prozesses beseitigen. Ein Null-Eins-Prozeß $\{x_n\}$, der in dieser Weise durch Vorgabe einer Lückenverteilung definiert ist, heiße *Lückenprozeß*.

Prozesse dieser Art sind von AKAIKE [1] beschrieben worden. Er hat gezeigt, daß sie immer eine absolut stetige Spektralfunktion besitzen. Hier wird gezeigt, daß die Spektraldichte eine rationale Funktion in $e^{2\pi i\lambda}$ ist, die explizit angegeben werden kann. Dies hat zur Folge, daß sich aus den Lückenprozessen durch geeignete gleitende Durchschnittsbildung über genau k aufeinanderfolgende Variable orthogonale Prozesse herleiten lassen.

In der Bezeichnung von FELLER [2] ist das Auftreten einer Eins in einem Lückenprozeß ein gewiß wiederkehrendes Ereignis mit einer endlichen mittleren Wartezeit L, für die

$$L = E\,\lambda = \sum_{i=1}^{k} i\,p_i = \frac{1}{p} \tag{1}$$

gilt.

1. Kovarianzfunktion

Die Kovarianzfunktion eines stationären Prozesses ist durch

$$R(n) = E\,\{(x_{m+n} - \mu)\,\overline{(x_m - \mu)}\}$$

definiert, worin $\mu = E\,x_n$ den von n unabhängigen Erwartungswert der x_n bedeutet. Für einen Null-Eins-Prozeß ergibt sich die einfache Schreibweise

$$R(n) = p\,\{x_{m+n} = 1,\, x_m = 1\} - p^2.$$

Wir führen noch die bedingten Wahrscheinlichkeiten

$$P_i = p\,\{x_{m+i} = 1 \,|\, x_m = 1\}$$

ein, die wegen der unabhängigen Lücken von m unabhängig sind. Dann ergibt sich die Kovarianzfunktion als

$$R(n) = p\,(P_n - p).$$

Speziell ist $P_0 = 1$ und $R(0) = p\,(1-p)$.

Für die bedingten Wahrscheinlichkeiten P_i bestehen Rekursionsformeln

$$P_n = p_1\,P_{n-1} + \cdots + p_n\,P_0 \qquad \text{für } n \leqq k \tag{2a}$$

und

$$P_n = p_1\,P_{n-1} + \cdots + p_k\,P_{n-k} \qquad \text{für } n \geqq k. \tag{2b}$$

Hieraus folgen Rekursionsformeln für die $R(n)$, es gilt

$$R(n) = p_1\,R(n-1) + \cdots + p_n\,R(0) - (1 - p_1 - \cdots - p_n)\,p^2 \qquad \text{für } n \leqq k \tag{3a}$$

und

$$R(n) = p_1\,R(n-1) + \cdots + p_k\,R(n-k) \qquad \text{für } n \geqq k. \tag{3b}$$

Die Relationen (3b) lassen sich auf eine lineare Abhängigkeit mit nur noch k Gliedern reduzieren.

SATZ 1. Es ist

$$R(n) + (1-p_1)\,R(n-1) + \cdots + (1 - p_1 - \cdots p_{k-1})\,R(n-k+1) = 0 \quad \text{für } n \geqq k-1.$$

BEWEIS. Die Beziehungen (3a) werden der Reihe nach für $n = 1$ bis $n = k-1$ hingeschrieben und addiert:

$$\begin{aligned}
&R(0) - p + p^2 = 0\\
&R(1) - p_1\,R(0) + (1-p_1)\,p^2 = 0\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
&R(k-1) - p_1\,R(k-2) - \cdots - p_{k-1}\,R(0) + (1 - p_1 - \cdots - p_{k-1})\,p^2 = 0.
\end{aligned}$$

Dabei ergeben die Glieder ohne ein $R(\nu)$ zusammen

$$p^2\big(k-(k-1)\,p_1-\cdots-(k-k)\,p_k\big)-p=p^2L-p=0,$$

was wegen (1) berschwindet. Die übrigen Glieder machen gerade die Behauptung für $n=k-1$ aus. Weiter schließt man durch vollständige Induktion mit (3b).

2. Charakteristisches Polynom

Es wird das Polynom

$$p(z)=z^k-p_1z^{k-1}-\cdots-p_k$$

gebildet, in dem nach Voraussetzung $p_k>0$ ist. Dieses Polynom hat $z_1=1$ als einfache Nullstelle, denn es ist gerade $p'(1)=L$, die mittlere Wartezeit nach (1), die immer positiv ist.

Definition. Das Polynom

$$p_0(z)=p(z)/(z-1)$$

heiße *charakteristisches Polynom* des Lückenprozesses mit der Lückenverteilung $p_1,\ldots,p_k$.

Es besitzt die Gestalt $p_0(z)=\sum_{j=0}^{k-1}B_jz^j$, worin

$$B_j=\sum_{i=k-j}^{k}p_i,$$

also speziell $B_{k-1}=1$ und $B_0=p_k>0$ ist. Auf das Polynom $p_0(z)$ läßt sich der Satz von Eneström-Kakeya [3] anwenden, der besagt, daß ein solches Polynom nur Nullstellen innerhalb des Einheitskreises besitzt. Voraussetzung für die Gültigkeit ist dabei noch die Bedingung, daß das Polynom $p(z)=(z-1)\,p_0(z)$ nicht als Polynom einer Potenz z^d, $d>1$ geschrieben werden kann. Diese Bedingung ist aber gerade durch die Forderung der Aperiodizität von $\{x_n\}$ gesichert. Die Nullstellen können mehrfach sein, was für die weitere Untersuchung keine Erschwerung bedeutet. Es gilt

Satz 2. Jedes Polynom

$$B(z)=\sum_{j=0}^{k-1}B_jz^j$$

mit

$$0<B_0\leqq B_1\leqq\cdots\leqq B_{k-1}=1$$

und $GGT(\nu_1,\ldots,\nu_\varkappa)=1$ für alle ν_η mit $B_{\nu_\eta}-B_{\nu_\eta-1}>0$ ist charakteristisches Polynom eines Lückenprozesses.

Beweis. Man setze

$$p_i=B_{k-i}-B_{k-i-1},\qquad i=1,\ldots,k.$$

Es besteht eine formale Übereinstimmung zwischen den Relationen (2b), (3b) und der Gleichung $p(z)=0$; man erhält die Relationen, wenn man z^n durch

P_n bzw. $R(n)$ ersetzt. Schließlich kann auch der Satz 1 noch in der Form

$$\sum_{j=0}^{k-1} B_j R(n+j) \qquad \text{für } n \geqq 0 \tag{4}$$

dargestellt werden.

Im folgenden wird die Beziehung $p_0(1) = \sum_{j=0}^{k-1} B_j = \frac{1}{p}$ benutzt.

Es werde nun zunächst vorausgesetzt, daß alle Nullstellen von $p_0(z)$ einfach sind. Dies bedeutet, daß es komplexe Zahlen $c_1, \ldots, c_k$ gibt, so daß die bedingten Wahrscheinlichkeiten P_n in der Form

$$P_n = c_1 z_1^n + \cdots + c_k z_k^n$$

geschrieben werden können. Dabei bedeuten $z_1, \ldots, z_k$ die Nullstellen von $p(z) = 0$, speziell sei $z_1 = 1$. Nach FELLER gilt $\lim_{n\to\infty} P_n = p$; daher kann $c_1 = p$ gesetzt werden und es folgt

$$P_n - p = \sum_{i=2}^{k} c_i z_i^n,$$

worin auf der rechten Seite nur noch Wurzeln z_i mit $|z_i| < 1$ erscheinen. Es ergibt sich daraus

$$\sum_{n=0}^{\infty} |R(n)| = p \sum_{n=0}^{\infty} |P_n - p| \leqq p \sum_{n=0}^{\infty} \sum_{i=2}^{k} |c_i|\,|z_i|^n < \infty. \tag{5}$$

Treten Nullstellen $z_1, \ldots, z_l$ mit der Ordnung $k_1, \ldots, k_l$ $\sum_{i=1}^{l} k_i = k$ auf, dann sind die c_i durch Polynome $c_i(n)$ zu ersetzen, die höchstens den Grad $k_i - 1$ in n haben können. Es ist dann entsprechend

$$P_n = c_1 z_1^n + c_2(n)\, z_2^n + \cdots + c_l(n)\, z_l^n,$$

worin c_1 wegen der Einfachheit von $z_1 = 1$ eine Konstante bleibt, für die wie vorher $c_1 = p$ gilt. Das Konvergenzverhalten der Reihe (5) wird durch die Mehrfachheit der Nullstellen nicht berührt. Die Reihe $\sum_{n=0}^{\infty} R(n)\, z^n$ besitzt einen Konvergenzradius $\varrho = \min_{i=2,\ldots,k} \frac{1}{|z_i|}$, $\varrho > 1$. Die absolute Konvergenz der Reihe (5) besagt, daß die zum Prozeß $\{x_n\}$ gehörende Spektralfunktion $W(\lambda)$ absolut stetig ist, also eine Spektraldichte $w(\lambda) = W'(\lambda)$ besitzt. Es ist

$$w(\lambda) = \sum_{n=-\infty}^{\infty} R(n)\, e^{2\pi i n \lambda}, \qquad R(n) = R(-n)$$

oder in reeller Schreibweise

$$w(\lambda) = R(0) + 2 \sum_{n=1}^{\infty} R(n) \cos 2\pi n \lambda.$$

3. Spektraldichte

Die Existenz einer Spektraldichte ist für Lückenprozesse von AKAIKE nachgewiesen worden. Es soll jetzt gezeigt werden, daß sich die Spektraldichte bei gegebener Lückenverteilung explizit berechnen läßt.

Satz 3. Es sei $\{x_n\}$ ein Lückenprozeß mit dem charakteristischen Polynom $\mathfrak{p}_0(z) = \sum_{j=0}^{k-1} B_j z^j$. Dann hat die Spektraldichte die Gestalt

$$w(\lambda) = \frac{H(e^{2\pi i\lambda})}{|\mathfrak{p}_0(e^{2\pi i\lambda})|^2},$$

worin

$$H(z) = \sum_{j=-k+2}^{k-2} H_j z^j \quad \text{mit} \quad H_j = H_{-j} = \mathfrak{p} \sum_{\nu=0}^{k-2-|j|} B_\nu (1 - B_{\nu+|j|}) \quad \text{für } j = 0, 1, \ldots, k-2.$$

Beweis. Es wird $H(z) = \mathfrak{p}_0(z)\, \mathfrak{p}_0\left(\frac{1}{z}\right) \sum_{n=-\infty}^{\infty} R(n)\, z^n$ angesetzt und daraus $H(z)$ berechnet.

Zunächst ist

$$\begin{aligned} S(z) &= \frac{1}{\mathfrak{p}}\, \mathfrak{p}_0(z) \sum_{n=0}^{\infty} R(n)\, z^{-n} \\ &= z^{k-1}[1 - B_{k-1}\mathfrak{p}] \\ &\quad + z^{k-2}[1 - (B_{k-1} + B_{k-2})\,\mathfrak{p}] \\ &\quad + \cdots \\ &\quad + z[1 - (B_{k-1} + \cdots + B_1)\,\mathfrak{p}]. \end{aligned}$$

Satz 1 bewirkt dabei das Verschwinden des absoluten Gliedes und der Koeffizienten der Glieder mit negativem Exponenten. Weiter ist jetzt

$$\frac{1}{\mathfrak{p}}\, H(z) = \mathfrak{p}_0\left(\frac{1}{z}\right) S(z) + \mathfrak{p}_0(z)\, S\left(\frac{1}{z}\right) - \frac{1}{\mathfrak{p}}\, R(0)\, \mathfrak{p}_0(z)\, \mathfrak{p}_0\left(\frac{1}{z}\right)$$

ein Polynom, in dem z nur noch mit den Exponenten j, $-(k-2) \leqq j \leqq k-2$ vorkommt. Glieder mit dem Exponenten $k-1$ kommen im ersten und letzten Term vor, mit dem Exponenten $-(k-1)$ im zweiten und letzten. Sie heben sich gegenseitig fort. Wegen der Symmetrie $w(\lambda) = w(-\lambda)$ ist auch $H_j = H_{-j}$. Eine einfache Ausrechnung bestätigt die angegebenen Werte. –

Die Spektraldichte läßt sich schließlich noch in die Form

$$w(\lambda) = \left| \frac{\sum_{j=0}^{k-2} A_j e^{2\pi i j\lambda}}{\sum_{j=0}^{k-1} B_j e^{2\pi i j\lambda}} \right|^2 \tag{6}$$

bringen, worin sich $A(z) = \sum_{j=0}^{k-2} A_j z^j$ aus $H(z) = A(z)\, A\left(\frac{1}{z}\right)$ ergibt.

Satz 3 bedeutet noch folgendes: Damit ein Null-Eins-Prozeß mit rationaler Spektraldichte in der Form (6) ein Lückenprozeß ist, muß $B(z) = \sum_{j=0}^{k-1} B_j z^j$ notwendig den Bedingungen des Satz 2 genügen, also charakteristisches Polynom eines Lückenprozesses sein. Dies ist aber im allgemeinen noch nicht hinreichend für das Vorliegen eines Lückenprozesses. Zu einem vorgegebenem charakteristischem Polynom $B(z)$ gibt es genau ein Polynom nach

der Ausrechnung im Beweise des Satz 3 und daraus genau ein Zählerpolynom $A(z)$. Erfüllt $A(z)$ diese Bedingungen, dann ist dies hinreichend für das Vorliegen eines Lückenprozesses.

4. Folgerungen, Beispiel

Besitzt ein stationärer Prozeß eine rationale Spektraldichte in der Form (6), dann läßt sich durch gleitende Durchschnittsbildung ein neuer Prozeß bilden, dessen Variable orthogonal sind. Es gilt

SATZ 4. Es sei $\{x_n\}$ ein Lückenprozeß mit dem charakteristischem Polynom $p_0(z) = \sum_{j=0}^{k-1} B_j z^j$ und $\{\xi_n\}$ durch

$$\xi_n = \sum_{j=0}^{k-1} B_j x_{n+j} - 1$$

definiert. Dann gilt

$$E\{\xi_m \xi_n\} = 0 \qquad \text{für } |m-n| \geqq k-1.$$

BEWEIS. Es handelt sich lediglich darum, den Satz 10.2 aus DOOB [4], Kap. X auf den hier vorliegenden Fall zu übertragen. Aus $E x_n = p$ und $\sum_{j=0}^{k-2} B_j = \frac{1}{p}$ folgt, daß $E\xi_n = 0$ ist.

Zum Schluß möge noch ein Beispiel mitgeteilt werden. Es sei $k=2$, dazu $p_1, p_2 \neq 0$ mit $p_1 + p_2 = 1$. Es muß $p_1 \neq 0$ sein, damit der Prozeß nicht periodisch wird. Dann ist $L = p_1 + 2p_2 = 1 + p_2$ und $p = \frac{1}{L} = \frac{1}{1+p_2}$. Das charakteristische Polynom ist $p_0(z) = z + p_2$. $H(z)$ wird zur Konstanten $H_0 = p\, B_0(1 - B_0) = p\, p_1 p_2$. Es folgt für die Spektraldichte

$$w(\lambda) = \frac{p_1 p_2}{1+p_2} \cdot \frac{1}{1 + p_2^2 + 2p_2 \cos 2\pi\lambda}.$$

Literatur

[1] AKAIKE, HIROTUGU: On a zero-one process and some of its applications. Ann. Inst. statist. Math. **8**, 87–94 (1956). — [2] FELLER, WILLIAM: An Introduction to Probability Theory and its Applications, Vol. I. New York 1952 (3. Printing). — [3] MARDEN, MORRIS: The Geometry of the zeros of a polynomial in a complex variable. New York 1949. — [4] DOOB, J. L.: Stochastic Processes. New York 1953.

Erlangen, Mathematisches Institut der Universität

(Eingegangen am 10. Januar 1959)

PETERSSON, H.
Math. Zeitschr. 71, 289–307 (1959)

Über Darstellungsanzahlen von Primzahlen durch Quadratsummen

Von
HANS PETERSSON

§ 1. Einleitung, methodische Hinweise, Ergebnisse

Die folgenden Ausführungen betreffen ein kürzlich diskutiertes Problem ([1] des Literaturverzeichnisses am Schluß der Abhandlung), die Zerlegung des Kreisteilungspolynoms

$$f(u) = \sum_{m=0}^{q-1} u^m = \sum_{m=0}^{q-1} \binom{q}{m+1} (u-1)^m \qquad (q = \text{Primzahl} > 3)$$

in zwei über dem quadratischen Zahlkörper $\mathsf{P}(\sqrt{\pm q})$ ($\pm q \equiv 1 \bmod 4$) irreduzible Faktoren. Bei Auszeichnung der primitiven q-ten Einheitswurzel $\exp \frac{2\pi i}{q}$ entsprechen diese den quadratischen Resten oder Nichtresten mod q, was durch die Schreibweise

$$f_\pm(u) = \sum_{m=0}^{r} \lambda_{m\pm} (u-1)^m \qquad (r = \tfrac{1}{2}(q-1),\ \lambda_{r\pm} = 1)$$

angedeutet werde. Das Hauptergebnis besteht in einem Zusammenhang der $\lambda_{m\pm}$ mit den q-ten Fourier-Koeffizienten $e_k(q)$ gewisser Eisensteinschen Reihen $E_k(\tau)$ für $k \equiv q \bmod 4$, $k \geqq 3$. Das einzelne $E_k(\tau)$ ist dadurch ausgezeichnet, daß es bei der analytischen Bestimmung der Darstellungsanzahlen als Summe von k Quadraten sowohl analytisch als auch arithmetisch den jeweiligen Hauptterm liefert.

Man gewinnt die übersichtlichste Interpretation dieses Zusammenhanges, indem man jeden Quotienten $\varrho_{m\pm} = \lambda_{m\pm} \lambda_{0\pm}^{-1}$ ($m = 1, 2, \ldots$) als Funktion der Primzahl q auffaßt ($q > 2m+1$). Nach Trennung der Fälle $q \equiv 1$ bzw. $3 \bmod 4$ läßt sich $\varrho_{m\pm}$ als Polynom in den $e_k(q)$ ($k \equiv q \bmod 4$, $3 \leqq k \leqq 2m+1$) mit explizit bestimmbaren Koeffizienten darstellen, die völlig elementar durch die Bernoullischen Zahlen B_{2n} ($2n \leqq m$) ausgedrückt und in jedem Einzelfall numerisch berechnet werden können. Für $q \equiv 3 \bmod 4$ ist dabei $e_3(q)$ durch die Klassenzahl h von $\mathsf{P}(\sqrt{-q})$ zu ersetzen.

Die konkrete Berechnung der $\varrho_{m\pm}$ für die niedrigsten Werte $m = 1, 2, 3$ hat gezeigt, daß sich diese Darstellungen mit wachsendem m schnell komplizieren, weshalb eine Fortsetzung der Berechnung kaum lohnend erscheint. Dagegen darf man Kongruenzen von der Art [1] (52) eines Interesses für wert

halten, *falls sie unmittelbar Darstellungsanzahlen von q durch Quadratsummen betreffen.* Für $k=7$ liegt dieser Sachverhalt nach [1] (52) vor: Hier ist $e_7(q)$ die Anzahl der Darstellungen von q als Summe von 7 Quadraten, die (für $q>7$) ganze Zahl e_7^* entsteht aus $e_7(q)$ durch Division durch 2520 bzw. 2072, je nachpem $q\equiv 3$ oder 7 mod 8 ist, und es gilt

$$e_7^* \equiv (q+1)\,h \bmod 3 \qquad (q\equiv 3 \bmod 4,\ q>7),$$

wo h wie oben die Klassenzahl von $\mathsf{P}(\sqrt{-q})$ bezeichnet. — Im folgenden sollen analoge Beziehungen aufgestellt werden.

Die Methode basiert auf dem in [1] angewendeten Schlußschema: In den erwähnten Darstellungen der $\varrho_{m_\pm}$ als Polynome in den $e_k(q)$ treten Nenner auf, die sich, weil die $\varrho_{m_\pm}$ ganze algebraische Zahlen sind, herausheben müssen. Als Bedingungen für diesen Sachverhalt ergeben sich Kongruenzen, denen gewisse der $e_k(q)$ nach den Primteilern der Nenner genügen. Die zusätzliche Forderung, daß in diesen Kongruenzen Darstellungsanzahlen von q durch Quadratsummen auftreten sollen, ist nun dahin zu präzisieren, daß die *tatsächlich auftretenden $e_k(q)$ selbst solche Darstellungsanzahlen sind.* Die Entscheidung hierüber verlangt ein genaues Eingehen auf das Problem der Bestimmung dieser Darstellungsanzahlen, aus dem die $e_k(q)$ entspringen.

Es sei $\tau=x+iy$ eine komplexe Variable mit $y=\operatorname{Im}\tau>0$, k eine ganze Zahl $\geqq 1$,

$$\vartheta(\tau)=\sum_{m=-\infty}^{+\infty} e^{\pi i m^2\tau}, \qquad \vartheta^k(\tau)=\sum_{n=0}^{\infty} a_k(n)\,e^{\pi i n\tau}, \tag{1.1}$$

so daß also $a_k(n)$ die Anzahl der Darstellungen von n als Summe von k Quadraten ganzer Zahlen bedeutet. Klassische Sätze erweisen $\vartheta^k(\tau)$ als ganze Modulform der Klasse $\left\{\Gamma_\vartheta, -\frac{k}{2}, v_\vartheta^k\right\}$, deren Bestimmungsstücke wie folgt definiert sind:

Γ_ϑ ist die von U^2 und T erzeugte Untergruppe der Gruppe Γ aller Modulmatrizen; stets sei

$$U=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \qquad T=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}. \tag{1.2}$$

Γ_ϑ ist eine Kongruenzgruppe der Stufe 2, hat in Γ den Index 3 und besitzt den Fundamentalbereich

$$\mathfrak{F}=\mathfrak{E}+U\,\mathfrak{E}+U\,T\,\mathfrak{E}, \tag{1.3}$$

wo $\mathfrak{E}$, das Grunddreieck der Modulfigur, aus den τ besteht, die das Gebiet $\{y>0,\ |x|<\frac{1}{2},\ |\tau|>1\}$ erfüllen, zuzüglich der Randpunkte τ dieses Gebietes mit $x\geqq 0$. $\mathfrak{F}$ hat genau zwei Spitzen ($\tau=\infty$ und $\tau=1$), einen einzigen elliptischen Fixpunkt ($\tau=i$; seine Ordnung ist $=2$) und nach vollzogener Ränderzuordnung das Geschlecht 0.

Das Multiplikatorsystem v_ϑ (gehörig zu Γ_ϑ und $-\frac{1}{2}$) ist in $\tau=\infty$ und $\tau=i$ unverzweigt und besitzt in der Spitze $\tau=1$ den Drehrest (Verzweigungs-

exponenten) $\frac{1}{8}$. Die explizite Bestimmung von v_ϑ gründet sich auf die definierende Eigenschaft von Γ_ϑ als der Gruppe der Modulmatrizen $L=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}$ mit

$$\beta \equiv \gamma \equiv 0 \bmod 2 \quad \text{oder} \quad \alpha \equiv \delta \equiv 0 \bmod 2; \tag{1.4}$$

dementsprechend ist

$$v_\vartheta(L)=\left(\frac{\gamma}{\delta}\right)_* \xi_4^{\delta-1} \quad \text{bzw.} \quad v_\vartheta(L)=\left(\frac{\delta}{\gamma}\right)^* \xi_4^{-\gamma}, \tag{1.5}$$

wo allgemein $\xi_m=\exp\frac{\pi i}{m}$ für ganzes $m>0$ und die beklammerten Brüche mit einem $*$ Jacobische Restsymbole bedeuten, die für negative Nenner so normiert sind, daß sie periodische Funktionen von δ darstellen.

Im folgenden bezeichnet $\mathfrak{R}^k$ die Klasse der Modulformen $\left\{\Gamma_\vartheta, -\frac{k}{2}, v_\vartheta^k\right\}$ (k ganz und $\geqq 1$). Versieht man die zur Spitze $\tau=\infty$ gehörige Eisenstein-Reihe dieser Klasse mit einem geeigneten konstanten Faktor, so erhält man das eingangs genannte

$$E_k(\tau)=1+\sum_{n=1}^{\infty} e_k(n)\, e^{\pi i n \tau} \in \mathfrak{R}^k \qquad (k \geqq 2). \tag{1.6}$$

Weil $\vartheta^k(\tau)$ in der Spitze $\tau=1$ die Ordnung $\frac{1}{8}k$ hat, gilt

$$\vartheta^k(\tau)=E_k(\tau)+f_k(\tau) \qquad (k \geqq 2) \tag{1.7}$$

mit einer ganzen Spitzenform f_k aus $\mathfrak{R}^k$. Man setze

$$f_k(\tau)=\sum_{n=1}^{\infty} b_k(n)\, e^{\pi i n \tau};$$

dann entspricht (1.7) die Zerlegung

$$a_k(n)=e_k(n)+b_k(n) \qquad (k \geqq 2,\ n \geqq 1). \tag{1.8}$$

Die Aussage, daß $e_k(n)$ eine Darstellungsanzahl sei, wird demgemäß durch

$$e_k(n)=a_k(n) \quad \text{oder} \quad b_k(n)=0 \tag{1.9}$$

präzisiert; diese Relation ist für ganze k, n mit $k \geqq 2, n \geqq 1$ in Betracht zu ziehen.

Über (1.9) weiß man folgendes: Für $2 \leqq k \leqq 8$ verschwindet $f_k(\tau)$ identisch, und daher gilt (1.9) für $1 \leqq k \leqq 8$ und alle $n \geqq 1$. Insbesondere sind also die (neben h) in [1] auftretenden $e_5(q), e_7(q)$ Darstellungsanzahlen. Für $k \geqq 9$ wurde in einigen wenigen Fällen (GLAISHER [2]) gezeigt, daß $f_k(\tau)$ nicht identisch verschwindet. Die vorliegende Untersuchung betrifft hauptsächlich die Fälle $k=9, 11$. Hier gilt (1.9) zwar nicht für alle $n \geqq 1$, wohl aber, wie im folgenden bewiesen wird, für die n einer gewissen (von k abhängigen) Folge, unter denen sich die Primzahlen $q \equiv k-4 \bmod 8$ befinden. Daher läßt sich $\varrho_{4\pm}$ für $q \equiv 5 \bmod 8$ als Polynom in den Darstellungsanzahlen $a_5(q), a_9(q)$, und es läßt sich $\varrho_{5\pm}$ für $q \equiv 7 \bmod 8$ als Polynom in h und den Darstellungsanzahlen $a_7(q), a_{11}(q)$ schreiben, und man kann dann das oben angedeutete Verfahren durchführen.

Für $k=11$ ist der hier erwähnte Sachverhalt nicht neu. Nach H. MAASS [3] (vgl. auch W. PFETZER [4]) stimmt $f_{11}(\tau)$ bis auf einen von Null verschiedenen konstanten Faktor mit

$$\vartheta_{3,4}(\tau)=\sum_{m_1,m_2,m_3=-\infty}^{+\infty}(m_1+i\,m_2)^4\,e^{\pi i\tau(m_1^2+m_2^2+m_3^2)} \tag{1.10}$$

überein, und daher verschwindet $b_{11}(n)$, wenn n nicht als Summe dreier Quadrate ganzer Zahlen darstellbar ist, also sicher für die n der Gestalt $4^a(8b+7)$ (a, b ganz und $\geqq 0$).

In Analogie zu (1.10) könnte man vermuten, daß $b_9(n)$ für alle n, die nicht Quadratzahlen sind, verschwindet, was auch im Hinblick auf die Ergebnisse von I. EBEL [5] über spezielle ternäre Darstellungsanzahlen nicht völlig abwegig ist. Diese Vermutung wird jedoch durch die Berechnung der $b_9(n)$ für die niedrigsten Werte von n widerlegt.

Um überhaupt *numerische* Aussagen über die $b_k(n)$ zu gewinnen, muß man ein *finites Konstruktionsverfahren* für die $f_k(\tau)$ aufstellen, das die numerische Berechnung der $b_k(n)$ für nicht zu große k und n mit einem vertretbaren Aufwand gestattet. Ein solches Verfahren wird in § 2 aus der divisorentheoretischen Methodik der Modulformen entwickelt. Von der praktischen Brauchbarkeit des Verfahrens habe ich mich unter anderem durch die Berechnung der $b_{17}(n)$ für $1\leqq n\leqq 16$ überzeugt. Im Nebenresultat ergibt sich ein kurzer Beweis eines Satzes von K. ANANDA RAU [6], der hier nicht nur für gerade, sondern für alle ganzen $k\geqq 9$ abgeleitet wird.

In § 3 werden *zwei Operatoren Heckescher Art* eingeführt, die für ganzes $k\geqq 1$, wenn $k\not\equiv 0 \bmod 8$, jeder ganzen Modulform $\varphi(\tau)$ aus der Klasse $\mathfrak{R}^k$ zwei ganze Modulformen der gleichen Dimension $-\frac{k}{2}$ zur Gruppe Γ^0 [16] zuordnen. Mit ihrer Hilfe bestätigen sich für $k=9, 10, 11$ die aus den Koeffiziententabellen zu vermutenden Gleichungen

$$b_k(n)=0 \quad \text{für } n\equiv k-4 \bmod 8,\quad n\geqq 1 \qquad (k=9,10,11), \tag{1.11}$$

und ein ähnlich gebauter Operator, den Herr K. WOHLFAHRT unabhängig von den oben erwähnten nach seinen allgemeinen Ansätzen [7] gebildet hat, ergibt

$$b_k(1)\,b_k(4n)=b_k(4)\,b_k(n) \quad \text{für alle } n\geqq 1 \qquad (k=9,10,11). \tag{1.12}$$

Im übrigen werden in § 3 Verallgemeinerungen der Relationen (1.11) auf die Werte $k=8g+r$ für $r=1, 2, 3$ und ganzes $g\geqq 2$ bewiesen, von denen der Sonderfall $k=17$ numerisch durchexerziert wird. Diese Ergebnisse zeigen, daß der operatorentheoretische Ansatz auch durch so allgemeine Untersuchungen wie [7] noch nicht entfernt ausgeschöpft wird.

In § 4 werden zunächst die reduzierten Koeffizienten $e_k^*=e_k^*(q)$ aus den $e_k(q)$ für $k=9, 11$ und $q\equiv k \bmod 4$ gebildet und mit den Methoden von [1] 5., 6. untersucht. Die Reduktion

$$e_k(q)=c_k^*(q)\,e_k^*(q) \quad (k\geqq 5,\ k\equiv 1 \bmod 2,\ q\equiv k \bmod 4)$$

erfolgt durch rationale Zahlen $c_k^*(q)$, die von q nur mod 8 abhängen, bei festem k also nur zwei Werte annehmen. Es war heuristisch nicht ohne Einfluß, daß die $c_k^*(q)$ für $k=9, 11$ genau dann ganz sind, wenn (1.9) mit $n=q$ zutrifft, d.h. für $q \equiv k-4 \bmod 8$.

Die sehr umständliche Durchführung des oben angedeuteten Verfahrens zur *Ableitung neuer Kongruenzen* aus der Diskussion der Darstellungen von $\varrho_{4\pm}$, $\varrho_{5\pm}$ ergibt unter anderem

(1.13) $e_{11}^* \equiv h \bmod 15$, wenn $q \equiv 7 \bmod 8$ und $q \equiv 7$ oder $13 \bmod 15$.

§ 2. Divisorentheorie und Konstruktionen in der Klasse $\mathfrak{R}^k$

2.1. Im folgenden werden Divisoren multiplikativ, und zwar als Potenzprodukte von (in Klammern gesetzten) Punkten des Fundamentalbereichs (1.3) $\mathfrak{F}$ von Γ_ϱ geschrieben. Demgemäß ist $(1)^{\frac{1}{2}k}$ der (genaue) Divisor von $\vartheta^k(\tau)$ $(k \geqq 1)$. Die Gesamtordnung des Divisors einer Form aus $\mathfrak{R}^k$ hat, wenn die Form nicht identisch verschwindet, den Wert $\frac{1}{8}k$. Da $E_k(\tau)$ in der Spitze $\tau=1$ von $\mathfrak{F}$ stets verschwindet, ist

(2.1) $$f_k(\tau) = \vartheta^k(\tau) - E_k(\tau) \equiv 0 \bmod (\infty)^1(1)^{\varepsilon_k} \qquad (k \geqq 2)$$

mit

(2.2) $$\varepsilon_k = \tfrac{1}{8}k - [\tfrac{1}{8}k], \quad \text{wenn} \quad k \not\equiv 0 \bmod 8, \quad \varepsilon_k = 1 \text{ sonst.}$$

Daraus folgt zunächst $f_k(\tau) \equiv 0$ für $2 \leqq k \leqq 8$.

Nach dem Riemann-Rochschen Satz ergibt sich für den Rang ν_k^+ der Schar $\mathfrak{C}_k^+$ der ganzen Spitzenformen aus $\mathfrak{R}^k$

(2.3) $$\nu_k^+ = \tfrac{1}{8}k - \varepsilon_k = \begin{cases} [\frac{1}{8}k], & \text{wenn } k \not\equiv 0 \bmod 8 \\ \frac{1}{8}k - 1, & \text{wenn } k \equiv 0 \bmod 8 \end{cases} \quad (k \geqq 1).$$

Für $9 \leqq k \leqq 16$ ist mithin eine Form aus $\mathfrak{C}_k^+$ bis auf einen konstanten Faktor eindeutig bestimmt. Wenn sie nicht identisch verschwindet, hat sie nach (2.1) den Divisor $(\infty)^1(1)^{\varepsilon_k}$.

Bildet man mit dem Legendreschen Modul $\varkappa^2$ die Funktion $z(\tau) = \varkappa^2(1-\varkappa^2)$ und benutzt zu ihrer Darstellung die ϑ-Nullwerte ϑ_3, ϑ_0, ϑ_2 in der Bezeichnung von Hurwitz $(\vartheta_3 = \vartheta(\tau))$, so wird

(2.4) $$z(\tau) = \vartheta_0^4 \vartheta_2^4 \vartheta_3^{-8}.$$

Aus den Produktentwicklungen dieser Nullwerte, insbesondere

$$\vartheta_2^4 = 16 u_2 \prod_{m=1}^{\infty} (1-u_1^m)^4 (1+u_1^m)^8, \qquad \vartheta_0^4 = \prod_{m=1}^{\infty} (1-u_1^m)^4 (1-u_2^{2m-1})^8$$

mit

$$u_2 = u_2(\tau) = e^{\pi i \tau}, \qquad u_1 = u_1(\tau) = e^{2\pi i \tau}$$

geht hervor, daß $z(\tau)$ in der oberen Halbebene holomorph ist und nicht verschwindet, bei der Transformation $\tau \to \tau + 2$ ungeändert bleibt und als Funktion von u_2 im Punkte $u_2 = 0$ die Ordnung 1 hat. Da sich überdies ϑ_0^4 und ϑ_2^4

bei der Transformation mit der Matrix T (abgesehen davon, daß sie beide den Faktor $-\tau^2$ aufnehmen) vertauschen, ist $z(\tau)$ eine vollinvariante, in der oberen Halbebene holomorphe und nicht verschwindende Funktion der Gruppe Γ_ϑ, die in $\tau=\infty$ die Ordnung $+1$, also in $\tau=1$ die Ordnung -1 und daher insgesamt den Divisor $(\infty)^1(1)^{-1}$ hat.

Andrerseits stellt $\Phi_k(\tau)=\vartheta^{-k}(\tau)\,f_k(\tau)$ $(k\geqq 9)$ ebenfalls eine vollinvariante Funktion der Gruppe Γ_ϑ dar mit

$$\Phi_k \equiv 0 \bmod (\infty)^1(1)^{-\nu_k^+} \qquad \left(\nu_k^+ \text{ nach } (2.3)\right).$$

Indem man den möglichen Pol von $\Phi_k(\tau)$ im Punkte $\tau=1$ mit Hilfe der Potenzen von $z(\tau)$ schrittweise abbaut, gelangt man zu der angekündigten Verallgemeinerung des Satzes von K. ANANDA RAU:

Satz 1. *Man definiere die Funktionen $f_k(\tau)$ durch* (1.1, 6, 7), *$\Phi_k(\tau)$ durch*

$$\Phi_k(\tau)=\vartheta^{-k}(\tau)\,f_k(\tau) \qquad (k\geqq 2).$$

Dann gilt: Für $2\leqq k\leqq 8$ verschwinden f_k und Φ_k identisch. Für $k\geqq 9$ läßt sich $\Phi_k(\tau)$ als ein Polynom $P_k(z)$ in der Funktion (2.4) $z=z(\tau)$ *darstellen. Der Grad dieses Polynoms P_k ist $\leqq \nu_k^+$ (vgl.* (2.3)) *und sein konstantes Glied verschwindet. Insbesondere ist $P_k(z)=\gamma_k z$ mit konstantem γ_k für $9\leqq k\leqq 16$.*

2.2. Wir konstruieren eine *Basis der Schar* $\mathfrak{C}_k^+$ für $k\geqq 9$. Es sei (vgl. (2.3))

$$r=r_k=8\varepsilon_k, \quad g=g_k=\nu_k^+, \quad \text{also} \quad k=8g+r, \quad 1\leqq r\leqq 8. \tag{2.5}$$

Die Formen φ aus $\mathfrak{C}_k^+$ sind durch $\varphi\in\mathfrak{R}^k$, $\varphi\equiv 0 \bmod (\infty)^1(1)^{\varepsilon_k}$ definiert, was soviel besagt wie

$$\vartheta^{-r}\varphi\in\mathfrak{R}^{8g}=\{\Gamma_\vartheta,\,-4g,\,1\}, \qquad \vartheta^{-r}\varphi\equiv 0 \bmod (\infty)^1.$$

Bezeichnet $\mathfrak{C}_{8g}^*$ die Schar der Multipla von $(\infty)^1$ in $\mathfrak{R}^{8g}$ $(g=1,2,\ldots)$ und $F_j^*(\tau)$ $(1\leqq j\leqq g)$ eine Basis von $\mathfrak{C}_{8g}^*$, so bilden die $\vartheta^r F_j^*$ eine Basis von $\mathfrak{C}_k^+$.

Es liegt nahe, die F_j^* multiplikativ aus Eisensteinschen Reihen der Klassen $\mathfrak{R}^{8m}$ $(1\leqq m\leqq g)$ zur Spitze $\tau=1$ zusammenzusetzen. Es ist $1=A^{-1}\infty$ mit $A=TU^{-1}=\begin{pmatrix}0 & -1\\ 1 & -1\end{pmatrix}$. Nach (1.4) hat die zweite Zeile $\underline{M}=\{m_1,m_2\}$ einer Matrix M aus $A\Gamma_\vartheta$ die Eigenschaft

$$(m_1,m_2)=1, \qquad m_1\equiv m_2\equiv 1 \bmod 2,$$

und umgekehrt gibt es zu solchen m_1, m_2 stets eine Matrix M in $A\Gamma_\vartheta$ mit $\underline{M}=\{m_1,m_2\}$. Wir bilden die fraglichen Reihen daher in der Gestalt

$$G_{-4m}(\tau)=\sum_{\substack{m_1,\,m_2=-\infty\\ m_1\equiv m_2\equiv 1\,(2)}}^{+\infty}(m_1\tau+m_2)^{-4m} \qquad (m \text{ ganz und } \geqq 1). \tag{2.6}$$

Alle diese G_{-4m} verhalten sich in der oberen τ-Halbebene holomorph und gestatten die Fourier-Entwicklungen

$$G_{-4m}(\tau)=-2\,\frac{\pi^{4m}}{(4m-1)!}\sum_{n=1}^{\infty}d_{4m-1}^*(n)\,e^{\pi i n\tau} \tag{2.7}$$

mit den Koeffizienten

$$d^*_{4m-1}(n) = (-1)^{n-1} \sum_{\substack{d|n,\, d>0 \\ n/d \equiv 1 \bmod 2}} d^{4m-1}, \qquad d^*_{4m-1}(1) = 1. \tag{2.8}$$

Das Bestehen der Funktionalgleichungen

$$G_{-4m}(L\tau) = (\gamma\tau + \delta)^{4m} G_{-4m}(\tau) \quad \text{für} \quad L = \begin{pmatrix} * & * \\ \gamma & \delta \end{pmatrix} \in \Gamma_\theta$$

bestätigt man direkt. Schließlich zeigt die Darstellbarkeit der Funktion

$$\tau^{-4m} G_{-4m}(A^{-1}\tau) = \sum_{\substack{m_1, m_2 = -\infty \\ m_1 \equiv 0,\, m_2 \equiv 1 \,(2)}}^{+\infty} (m_1\tau + m_2)^{-4m}$$

durch eine in der oberen Halbebene holomorphe Potenzreihe in der Variablen $e^{2\pi i\tau}$ mit von Null verschiedenem konstantem Glied, daß $G_{-4m} \in \mathfrak{G}^*_{8m}$. Im Punkte $\tau = \infty$ hat G_{-4m} die Ordnung 1, in $\tau = 1$ die Ordnung 0.

Wir setzen

$$F_{-4m}(\tau) = -\tfrac{1}{2}(4m-1)!\,\pi^{-4m} G_{-4m}(\tau) = \sum_{n=1}^{\infty} d^*_{4m-1}(n)\, e^{\pi i n\tau} \tag{2.9}$$

und erhalten

Satz 2. *Die Modulformen* $F^j_{-4} F_{-4(g-j)}$ $(0 \leqq j \leqq g-1$; *vgl.* (2.5, 6, 7, 9)) *bilden eine Basis von* $\mathfrak{G}^*_{8g}$ $(g \geqq 1)$, *und demgemäß bilden die Modulformen*

$$H^{(r)}_{g,j} = \vartheta^r F^j_{-4} F_{-4(g-j)} \qquad (0 \leqq j \leqq g-1)$$

eine Basis von $\mathfrak{G}^+_k$ $(k \geqq 9)$. *Für* $9 \leqq k \leqq 16$ *ist* $g = 1$, *und die Schar* $\mathfrak{G}^+_k$ *wird von der einen Funktion* $H^{(r)}_{1,0} = \vartheta^r F_{-4}$ *aufgespannt.*

2.3. Diese Konstruktion ist für $k = 9, 17$ wie folgt zu realisieren: *Im Falle* $k = 9$ hat man

$$f_9(\tau) = C_9 H^{(1)}_{1,0}(\tau) = C_9 \sum_{n=1}^{\infty} c_9(n)\, e^{\pi i n\tau} \tag{2.10}$$

mit

$$c_9(n) = d^*_3(n) + 2 \sum_{1 \leqq l < \sqrt{n}} d^*_3(n - l^2) \qquad (n \geqq 1) \tag{2.11}$$

und $C_9 = a_9(1) - e_9(1)$; die Berechnung ergibt auf Grund von [3]: $C_9 = \frac{3}{17}$.

Der Fall $k = 17$ ist numerisch viel mühsamer. Hier ist $g = 2$ und daher besteht nach Satz 2 eine Basis von $\mathfrak{G}^*_{8\cdot 2} = \mathfrak{G}^*_{16}$ aus den Funktionen

$$F_{-8}(\tau) = \sum_{n=1}^{\infty} d^*_7(n)\, e^{\pi i n\tau}, \qquad F^2_{-4}(\tau) = \sum_{n=2}^{\infty} d^*_{3,2}(n)\, e^{\pi i n\tau}$$

mit

$$d^*_{3,2}(n) = \sum_{h=1}^{n-1} d^*_3(h)\, d^*_3(n-h) \qquad (n \geqq 2).$$

Daher gewinnt man eine Basis von $\mathfrak{G}^+_{17}$ in den beiden Funktionen

$$H^{(1)}_{2,0}(\tau) = \sum_{n=1}^{\infty} c_{17}(n)\, e^{\pi i n\tau}, \qquad H^{(1)}_{2,1}(\tau) = \sum_{n=2}^{\infty} c'_{17}(n)\, e^{\pi i n\tau},$$

mit

$$c_{17}(n) = d_7^*(n) + 2\sum_{1\le l<\sqrt{n}} d_7^*(n-l^2),$$

$$c'_{17}(n) = d_{3,2}^*(n) + 2\sum_{1\le l<\sqrt{n}} d_{3,2}^*(n-l^2),$$

und nun gilt

$$\vartheta^{17}(\tau) = E_{17}(\tau) + C_{17}H_{2,0}^{(1)}(\tau) + C'_{17}H_{2,1}^{(1)}(\tau)$$

mit gewissen numerischen Konstanten C_{17}, C'_{17}. In C'_{17} geht nach [3] der Wert der Dirichletschen Reihe $L(8,\chi)$ für den geraden Charakter $\chi(m) = \left(\frac{2}{m}\right) \bmod 8$ ein. Man bestimmt diesen Wert aus [1] (36) mit dem Ergebnis

$$L(8,\chi) = \pi^8 \frac{24611\cdot\sqrt{2}}{2^{20}\cdot 3^2\cdot 5\cdot 7} \sim (1{,}0001545)^{-1}$$

und findet danach

$$C_{17} = \frac{2^8\cdot 7\cdot 17\cdot 4003}{257\cdot 3617}, \quad C'_{17} = \frac{2^{11}\cdot 2024513}{257\cdot 3617}.$$

Für die *Abschätzungen der Fourier-Koeffizienten* der ganzen Spitzenformen aus der Klasse $\mathfrak{R}^k$ sind die Basen des Satzes 2 wertlos. Ein Idealfall, der die finite Darstellbarkeit der Fourier-Koeffizienten mit der Möglichkeit, sie optimal abzuschätzen, vereinigt, ist der Fall $k = 10$. Hier hat man

$$e_{10}(n) = \frac{4}{5}\sum_{d\mid n,\, d>0} \left(\frac{-4}{d}\right)\left\{\left(\frac{2n}{d}\right)^4 + d^4\right\} \qquad (n \ge 1),$$

$$\vartheta^2(\tau)F_{-4}(\tau) = \tfrac{1}{4}\vartheta_{2,4}(\tau) = \tfrac{1}{4}\sum_{m_1,m_2=-\infty}^{+\infty} (m_1 + i\,m_2)^4 e^{\pi i(m_1^2+m_2^2)\tau},$$

$$f_{10}(\tau) = \sum_{n=1}^{\infty} b_{10}(n)\, e^{\pi i n\tau} = \tfrac{8}{5}\vartheta_{2,4}(\tau), \tag{2.12}$$

und daher wird

$$b_{10}(n) = \tfrac{8}{5}\sum_{m_1^2+m_2^2=n} (m_1^4 - 6m_1^2m_2^2 + m_2^4) = \tfrac{8}{5}\sum_{m_1^2+m_2^2=n} (n^2 - 8m_1^2m_2^2), \tag{2.13}$$

insbesondere aber $b_{10}(n) = 0$, wenn n nicht als Summe zweier Quadrate dargestellt werden kann. Dies gilt speziell für die Primzahlen $p \equiv 3 \bmod 4$, während für eine Primzahl $p \equiv 1 \bmod 4$

$$-p^2 + 4p - 2 \le \tfrac{5}{64} b_{10}(p) \le p^2 - 8(p-1)$$

resultiert. Durchläuft n die natürlichen Zahlen, so folgt aus (2.13)

$$b_{10}(n) = O(n^{2+\varepsilon}) \quad \text{für } n \to \infty \text{ und beliebiges } \varepsilon > 0.$$

Für $k = 11$ wird in [3]

$$f_k(\tau) = f_{11}(\tau) = \tfrac{88}{31}\vartheta_{3,4}(\tau) \tag{2.14}$$

bestimmt; zur Bezeichnung vgl. (1.10). Ersichtlich ist $\vartheta_{3,4}(\tau) = \vartheta(\tau)\vartheta_{2,4}(\tau)$.

§ 3. Spezielle Operatoren Heckescher Art

Bezeichnet $f(\tau)$ eine Modulform, die als solche zu einer Untergruppe Γ_ϱ von endlichem Index in der Modulgruppe Γ und zu einer reellen Dimension $-\varrho$ gehört, hat ferner die Matrix $S=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ reelle Elemente a, b, c, d und eine positive Determinante, so schreiben wir

$$(3.1) \qquad f|S = f(\tau)|S = f(S\tau)(c\tau+d)^{-\varrho}\left(S\tau = S(\tau) = \frac{a\tau+b}{c\tau+d}\right).$$

Wir untersuchen zunächst die Wirkung der Transformation $\tau \to N^{-1}\tau$, wo N eine feste natürliche Zahl angibt. Es sei $L=\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \Gamma$, $L_N^* = \begin{pmatrix} \alpha & N^{-1}\beta \\ N\gamma & \delta \end{pmatrix}$, $\mathfrak{C}_k$ die Schar der ganzen Modulformen aus der Klasse $\mathfrak{R}^k$ (k ganz und $\geqq 1$), $\varphi(\tau)$ eine Form aus $\mathfrak{C}_k$, $\psi_N(\tau)=\varphi(N^{-1}\tau)$. Wegen $N^{-1}L(\tau)=L_N^*(N^{-1}\tau)$ gilt, wenn L_N^* in Γ_ϱ liegt:

$$(3.2) \qquad \psi_N(\tau)|L = v_\varrho^k(L_N^*)\,\psi_N(\tau).$$

Wir nehmen $N \equiv 0 \bmod 2$. Die Bedingung $L_N^* \in \Gamma_\varrho$ besagt für $L \in \Gamma$ genau, daß $\beta \equiv 0 \bmod 2N$, was in der traditionellen Bezeichnung durch $L \in \Gamma^0[2N]$ auszudrücken ist. $\psi_N(\tau)$ stellt also eine ganze Modulform der Klasse

$$(3.3) \qquad \mathfrak{R}_*^k[N] = \{\Gamma^0[2N], -\tfrac{1}{2}k, v_{*\varrho}^k\} \qquad (N \equiv 0 \bmod 2)$$

dar, wo $v_{*\varrho}$ nach (1.5), (3.2) erklärt ist durch

$$(3.4) \qquad v_{*\varrho}(L) = \left(\frac{N\gamma}{\delta}\right)_* \xi_4^{\delta-1} \quad \text{für} \quad L=\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \Gamma^0[2N].$$

Es sei $l = l_k$ eine ganze Zahl, $\varphi(\tau) = \sum_{n=0}^{\infty} \beta_n e^{\pi i n \tau}$. Dann hat

$$(3.5) \qquad F_{kl,N}(\tau) = N^{-1} \sum_{j \bmod N} \xi_N^{-2klj}\,\psi_N(\tau + 2j)$$

die Fourier-Entwicklung

$$(3.6) \qquad F_{kl,N}(\tau) = \sum_{\substack{n=0 \\ n \equiv kl \bmod N}}^{\infty} \beta_n e^{\pi i n \frac{\tau}{N}}.$$

Wir setzen $N = 2^\nu$ (ν ganz) und versuchen, l so zu bestimmen, daß $F_{kl,N}$ eine Modulform der Klasse $\mathfrak{R}_*^k[N]$ wird. Bei gegebenem $L \in \Gamma^0[2N]$ führt dies auf die Frage nach der Zuordnung der Restklassen j und $j' \bmod N$ durch die Bedingung, daß

$$(3.7) \qquad \widetilde{L}_{j,j'} = U^{2j} L\, U^{-2j'} = \begin{pmatrix} \alpha + 2\gamma j & \beta + 2\delta j - 2\alpha j' - 4\gamma j j' \\ \gamma & \delta - 2\gamma j' \end{pmatrix}$$

in $\Gamma^0[2N]$ enthalten sei, was soviel besagt wie

$$(3.8) \qquad (\alpha + 2\gamma j)\, j' \equiv \delta j \bmod 2^\nu.$$

Die Zuordnung ist also eineindeutig und erfüllt $(j, 2^r) = (j', 2^r)$. Nach (3.1, 2, 4, 5) ergibt sich

$$F_{kl,N}(\tau)\,|\,L = N^{-1} \sum_{j \bmod N} \xi_N^{-2klj} v_*^k \left(\tilde{L}_{j,j'}\right) \psi_N(\tau + 2j'),$$

und daher stellt $F_{kl,N}$ eine ganze Modulform aus $\mathfrak{R}_*^k[N]$ dar, wenn durch Wahl eines nur von k abhängigen l folgendes erreicht werden kann:

Für jedes $L \in \Gamma^0[2N]$ und jedes ganze j gilt

$$\xi_N^{2klj'} \left(\frac{N\gamma}{\delta - 2\gamma j'}\right)_*^k \xi_4^{-2k\gamma j'} = \xi_N^{2klj} \left(\frac{N\gamma}{\delta}\right)_*^k \tag{3.9}$$

mit einem ganzen j', das der Kongruenz (3.8) genügt.

Wir nehmen definitiv $N = 8$. Dann hat man

$$\xi_4^{k(l-2\gamma)j'} \left(\frac{2\gamma}{\delta - 2\gamma j'}\right)_*^k = \xi_4^{klj} \left(\frac{2\gamma}{\delta}\right)_*^k \tag{3.10}$$

unter den Bedingungen

$$(1 + 2\gamma\delta j)\, j' \equiv j \bmod 8 \quad \text{und} \quad (2\gamma, \delta) = 1 \tag{3.11}$$

zu erfüllen. Ersichtlich kommt l wie k nur mod 8 in Betracht.

Als Äquivalent dieser Aussage ergibt sich für gerades k leicht

$$l \equiv 1 \bmod 2 \quad \text{für} \quad k \equiv 2 \bmod 4; \quad l \text{ beliebig ganz für } k \equiv 0 \bmod 4.$$

Es sei $k \equiv 1 \bmod 2$. In der Aussage kann man, ohne sie zu ändern, $k = 1$ setzen. Für $\gamma \equiv 0 \bmod 4$ ist $j' \equiv j \bmod 8$, und die Aussage gilt für beliebiges ganzes l. Im Falle $\gamma \equiv 2 \bmod 4$ stimmen die Restsymbole beiderseits überein, und für $j' \equiv j \equiv 0 \bmod 2$ ist wieder $j' \equiv j \bmod 8$; die Aussage gilt dann auch für beliebiges ganzes l. Wenn aber $j' \equiv j \equiv 1 \bmod 2$ zutrifft, so ist $j' \equiv j + 4 \bmod 8$ und die Aussage nun mit $l \equiv 1 \bmod 2$ gleichbedeutend. Daher werde jetzt $l \equiv 1 \bmod 2$ vorausgesetzt.

Im Falle $\gamma \equiv 1 \bmod 2$ wende man auf das Restsymbol der linken Seite zweimal das quadratische Reziprozitätsgesetz an. Dann ergibt sich nach (3.11) als Äquivalent der Aussage die unter der Bedingung (3.11) zu erfüllende Relation

$$i^w = 1 \quad \text{mit} \quad w = -l\gamma\delta jj' - j' + (\gamma j' - \delta) j',$$

die für $j \equiv j' \equiv 0 \bmod 2$ ersichtlich zutrifft. Für $j \equiv j' \equiv 1 \bmod 2$ schließt man aus (3.11) zunächst $j' \equiv -j \bmod 4$ und daraus

$$w \equiv l\gamma\delta + j + \gamma + \delta j = (\gamma + \delta j)(1 + l\delta) - j(l - 1) \bmod 4.$$

Zusammenfassend findet man also: Für $k \equiv 1 \bmod 2$ bedeutet (3.10) unter der Bedingung (3.11) soviel wie $l \equiv 1 \bmod 4$.

Wir definieren jetzt für beliebiges ganzes m die folgenden auf die ganzen Formen φ aus $\mathfrak{R}^k$ anzuwendenden *Operatoren* Y_m (vgl. (3.5, 6)):

$$\varphi\,|\,Y_m = \varphi(\tau)\,|\,Y_m = 2^{-3} \sum_{j \bmod 8} \xi_4^{-mj} \varphi\left(\frac{\tau + 2j}{8}\right) = \sum_{\substack{n=0 \\ n \equiv m \bmod 8}}^{\infty} \beta_n e^{\pi i n \frac{\tau}{8}}. \tag{3.12}$$

Y_m hängt ersichtlich nur von der Restklasse von m mod 8 ab. – Es gilt

Satz 3. *Es sei $\varphi(\tau)$ eine ganze Modulform aus der Klasse $\mathfrak{R}^k$ $(k \geqq 1)$. Die Funktion* (3.13) $\varphi(\tau)|\, Y_m$ *stellt eine ganze Modulform aus der Klasse* (3.3) $\mathfrak{R}^k_* [8] = \{\Gamma^0[16], -\frac{1}{2}k, v^k_{*\,\delta}\}$ *dar, wenn m die in der folgenden Tabelle angegebenen Werte hat; dabei ist $r \equiv k \bmod 8$, $1 \leqq r \leqq 8$:*

r	1	2	3	4	5	6	7	8
m	1,5	2,6	3,7	0,4	1,5	2,6	3,7	0

Für einen Teil der Anwendungen dieses Satzes bedarf man genauerer Kenntnisse über die Klassen $\mathfrak{R}^k_* [8]$. Zunächst sind die *Äquivalenzklassen der Spitzen nach der* $\Gamma^0[16]$ zu bestimmen. Diese entnimmt man dem folgenden allgemeinen Satz:

Für ganzes $N > 1$ bestehe $\Gamma^0[N]$ aus den Modulmatrizen $L = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ mit $\beta \equiv 0 \bmod N$. Es sei q eine Primzahl $\geqq 2$, ν eine natürliche Zahl. Ein volles Vertretersystem der Spitzen mod $\Gamma^0[q^\nu]$ *wird (mit den zugehörigen Breiten) gegeben durch*

$$\infty\,(\textit{Breite}\ q^\nu); \qquad 0\,(\textit{Breite}\ 1);$$

$$h q^\lambda \quad \textit{mit} \quad \tfrac{1}{2}\nu \leqq \lambda \leqq \nu - 1, \quad h \bmod q^{\nu-\lambda}, \quad (h, q) = 1 \qquad (\textit{Breite}\ 1)$$

$$h q^\lambda \quad \textit{mit} \quad 1 \leqq \lambda \leqq \tfrac{1}{2}\nu, \qquad\quad h \bmod q^{\lambda}, \quad (h, q) = 1 \qquad (\textit{Breite}\ q^{\nu-2\lambda}).$$

Im vorliegenden Falle ist $q = 2$, $\nu = 4$. Man erhält ein volles Vertretersystem der Spitzen mod $\Gamma^0[16]$ in den Punkten (mit den Breiten)

$$\infty(16),\ 0(1),\ 8(1),\ 4(1),\ -4(1),\ 2(4),$$

und man stellt danach leicht fest, daß die Drehreste $\varkappa$ des Multiplikatorsystems $v_{*\,\delta}$ in allen Spitzen der $\Gamma^0[16]$ verschwinden.

Ferner gilt $\Gamma^0[16] = D_2 \Gamma^0_0[4] D_2^{-1}$ mit $D_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2^{-1} \end{pmatrix}$ und der traditionellen Bezeichnung $\Gamma^0_0[N]$ für die Gruppe der Modulmatrizen $L = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$, welche $\beta \equiv \gamma \equiv 0 \bmod N$ erfüllen. Daraus folgt sofort, daß $\Gamma^0[16]$ das Geschlecht Null hat. Da elliptische Fixpunkte offenbar nicht vorhanden sind, liefert der Riemann-Rochsche Satz für die Ränge $\nu_{*\,k}$ bzw. $\nu^+_{*\,k}$ der Scharen $\mathfrak{C}_{*\,k}$ bzw. $\mathfrak{C}^+_{*\,k}$ der ganzen Formen bzw. Spitzenformen aus $\mathfrak{R}^k_* [8]$

$$\nu_{*\,k} = k + 1, \qquad \nu^+_{*\,k} = k - 5 \qquad (k \geqq 5). \tag{3.13}$$

Die Ordnung des Divisors einer Form aus $\mathfrak{R}^k_* [8]$ ist, wenn die Form nicht identisch verschwindet, gleich k $(k \geqq 1)$. Man erhält daher den folgenden

Satz 4. *Es bezeichne $\mathfrak{C}_{*\,k}$ die Schar der ganzen Formen, $\mathfrak{C}^+_{*\,k}$ die Schar der ganzen Spitzenformen aus der Klasse $\mathfrak{R}^k_* [8]$, $\nu_{*\,k}$ den Rang von $\mathfrak{C}_{*\,k}$, $\nu^+_{*\,k}$ den Rang von $\mathfrak{C}^+_{*\,k}$. Dann gilt* (3.13) *für $k \geqq 5$. Es sei $k \geqq 1$. Eine Modulform*

$$\varphi_*(\tau) = \sum_{n=0}^{\infty} \beta_{*\,n} e^{\pi i n \frac{\tau}{8}} \in \mathfrak{C}_{*\,k}$$

verschwindet identisch, wenn $\beta_{*n}=0$ *für* $0\leqq n\leqq k$. *Liegt* $\varphi_*(\tau)$ *sogar in* $\mathfrak{C}^*_{*k}$, *so verschwindet* $\varphi_*(\tau)$ *sowohl dann identisch, wenn* $k\leqq 5$ *ist, als auch dann, wenn* $k\geqq 6$ *und* $\beta_{*n}=0$ *für* $1\leqq n\leqq k-5$ *zutrifft.*

In den ersten Anwendungen von Satz 3 tritt die durch

$$\vartheta_{2*}(\tau)=\vartheta_2\Big(\frac{\tau}{2}\Big),\qquad \vartheta_2(\tau)=\sum_{m=-\infty}^{+\infty} e^{\pi i(m+\frac{1}{2})^2\tau}$$

erklärte Modulform $\vartheta_{2*}(\tau)$ auf. Bekanntlich gehört $\vartheta_2(\tau)$ zur Gruppe $\Gamma_0[2]$ (definiert durch $\gamma\equiv 0 \bmod 2$), woraus sofort folgt, daß $\vartheta_{2*}(\tau)$ zur $\Gamma^0[2]$ und daher auch zur $\Gamma^0[16]$ gehört. Durch Koeffizientenvergleich ergibt sich nach (3.12) unmittelbar

Satz 5. *Im Sinne von Satz 3 ist*

$$\vartheta(\tau)|Y_1=\vartheta_{2*}(\tau),\qquad \vartheta^2(\tau)|Y_2=\vartheta^2_{2*}(\tau),\qquad \vartheta^3(\tau)|Y_3=\vartheta^3_{2*}(\tau),$$
$$\vartheta(\tau)|Y_5=\vartheta^2(\tau)|Y_6=\vartheta^3(\tau)|Y_7=0.$$

Wir betrachten jetzt die *ganzen Spitzenformen der Klassen* $\mathfrak{R}^k$ *für* $k=9$, 10, 11; zu jedem dieser k existiert eine und, von einem konstanten Faktor abgesehen, nur eine solche Spitzenform $\varphi_k(\tau)$, die nicht identisch verschwindet, nach Satz 2 etwa

$$\varphi_k(\tau)=H^{(r)}_{1,0}(\tau)=\sum_{n=1}^{\infty} c_k(n)\,e^{\pi i n\tau}\qquad (r=1,2,3).$$

Wir wollen zeigen, daß $Y_{k-4}=Y_{r+4}$ dieses φ_k in Null überführt. Der Beweis kann auf zwei Arten geführt werden. Benutzt man, daß $\varphi_k|Y_{k-4}$ eine ganze Spitzenform der Klasse $\mathfrak{R}^k_*$ [8] ist und daß die Drehreste dieser Klasse sämtlich verschwinden, so zählt man ab, daß $\varphi_k|Y_{k-4}$ insgesamt mindestens $k+1$ Nullstellen besitzt, also identisch verschwindet. Ein zweiter Beweis, der die etwas mühsame Bestimmung der Spitzen, Spitzenbreiten und Drehreste umgeht, stützt sich lediglich auf den Satz von der Gesamtordnung des Divisors einer nicht identisch verschwindenden Modulform. Zur Anwendung dieses Satzes braucht man nur zu wissen, daß $\Gamma^0[16]$ in Γ den Index 24 hat. Nach (2.11) errechnet man $c_9(5)=b_9(5)=0$. Auf die gleiche Weise kann man

$$c_{10}(6)=b_{10}(6)=c_{11}(7)=b_{11}(7)=0$$

bestätigen, doch ergibt sich dies auch unmittelbar aus (2.13, 14) und (1.10). Nun folgt, daß $\varphi_k|Y_{k-4}$ mindestens $k+4$ Nullstellen besitzt, also identisch verschwindet.

Satz 6. *Für* $k=9$, 10, 11 *gilt* $f_k(\tau)|Y_{k-4}=0$. *Es ist also* (*vgl.* (1.7, 8))

$$b_k(n)=0,\quad \textit{wenn}\quad n\equiv k-4 \bmod 8,\quad n>0.$$

Eine Verallgemeinerung dieses Satzes besteht für die Werte $k=8g+r$ mit $g\geqq 2$, $r=1,2,3$. Nach Satz 4 und (3.13) weiß man: Zu jedem System beliebig gegebener komplexer Zahlen c_{*n} $(1\leqq n\leqq k-5)$ existiert genau eine Form

$\varphi_*(\tau)$ in $\mathfrak{C}^+_{*k}$ mit $\beta_{*n}=c_{*n}$ $(1\leqq n\leqq k-5)$. Daher gibt es genau $g-1$ linear unabhängige Formen

$$\psi_*(\tau)=\sum_{n=1}^{\infty}\gamma_{*n}e^{\pi i n\frac{\tau}{8}}\in\mathfrak{C}^+_{*k},$$

für die $\gamma_{*n}=0$ ist, wenn $n\equiv r+4 \bmod 8$, $1\leqq n\leqq k-5$. Mithin gilt

Satz 7. *Es sei $k=8g+r$, $g\geqq 2$, $r=1, 2, 3$. Zwischen den g Modulformen*

$$H^{(r)}_{g,j}(\tau)\,|\,Y_{r+4}\in\mathfrak{C}^+_{*k}\qquad (j=0,1,2,\ldots,g-1)$$

besteht eine nicht-triviale lineare Relation mit konstanten Koeffizienten. Das lineare Vektorgebilde der Koeffizienten-Vektoren aller dieser Relationen besitzt eine Basis, deren Vektoren lauter ganze rationale Komponenten haben.

Wir geben von diesem Satz eine Anwendung auf den konkreten Fall

$$k=17\qquad (g=2,\ r=1;\ \text{vgl. § 2.3}).$$

Die numerische Rechnung liefert die Tabelle

n	1	2	5	13
$a_{17}(n)$	34	544	199104	
$257\cdot 3617\,e_{17}(n)$	$2\cdot 17\cdot 67\cdot 491$	$2^5\cdot 3\cdot 5\cdot 17\cdot 24611$	$2^8\cdot 3^2\cdot 5\cdot 17\cdot 19^2\cdot 43\cdot 257$	
$c_{17}(n)$	1	-126	45360	589680
$c'_{17}(n)$	0	1	-336	-4368

und damit den folgenden

Satz 8. (*Bezeichnungen nach* § 2). *Für die natürlichen Zahlen $n\equiv 5 \bmod 8$ gilt $c_{17}(n)+135\,c'_{17}(n)=0$, d.h. in der Bezeichnung* (2.8)

$$d^*_7(n)+2\sum_{1\leqq l<\sqrt{n}}d^*_7(n-l^2)+135\left\{d^*_{3,2}(n)+2\sum_{1\leqq l<\sqrt{n}}d^*_{3,2}(n-l^2)\right\}=0$$

und $a_{17}(n)=e_{17}(n)-C^*_{17}c_{17}(n)$, *d.h.*

$$a_{17}(n)=e_{17}(n)-C^*_{17}\left\{d^*_7(n)+2\sum_{1\leqq l<\sqrt{n}}d^*_7(n-l^2)\right\}$$

mit

$$C^*_{17}=\frac{2^8\cdot 476221}{3^3\cdot 5\cdot 257\cdot 3617}=\frac{2^8\cdot 17\cdot 109}{3^3\cdot 5\cdot 3617}.$$

Eine Prüfung der Rechnung ergab sich auf folgende Weise: Der Proportionalitätsfaktor 135 in der ersten Gleichung von Satz 8, zu dessen Bestimmung es nach Satz 7 nur der Werte von $c_{17}(n)$, $c'_{17}(n)$ für $n=5$ bedarf, bestätigte sich an den Werten $c_{17}(13)$, $c'_{17}(13)$. Die Konstanten C_{17}, C'_{17} aus § 2, 3 sind durch die Tabellenwerte für $n=1, 2$ zu ermitteln. Die entscheidende Probe bestand in der Bestätigung der aus der Theorie leicht abzuleitenden Gleichung

$$a_{17}(5)-e_{17}(5)=336\,(135\,C_{17}-C'_{17}).$$

Zum Schluß soll noch das eingangs erwähnte Ergebnis von K. WOHLFAHRT abgeleitet werden. Der von ihm eingeführte Operator läßt sich in dem hier entwickelten Formalismus durch

$$(3.14)\qquad \varphi(\tau)|Z_4 = \tfrac{1}{4}\sum_{j \bmod 4} \psi_4(\tau+2j) = F_{0,4}(\tau) \qquad (N=4,\ l=0)$$

erklären ($\varphi \in \mathfrak{G}_k$, $k \geqq 1$). Für $L \in \Gamma_\vartheta \cap \Gamma^0[8]$ wird $j' \equiv j \bmod 4$, und (3.9) geht in die leicht beweisbare Relation

$$\left(\frac{\gamma}{\delta - 2\gamma j'}\right)_*^k \xi_4^{-2k\gamma j'} = \left(\frac{\gamma}{\delta}\right)_*^k \qquad (L \in \Gamma_\vartheta \cap \Gamma^0[8])$$

über. Daraus folgt

$$\varphi(\tau)|Z_4 \in \{\Gamma_\vartheta \cap \Gamma^0[8],\ -\tfrac{1}{2}k,\ v_\vartheta^k\}.$$

Außerdem aber ist $F_{0,4}(\tau+2) = F_{0,4}(\tau)$; da man zu jedem L_1 aus (der Hauptkongruenzgruppe 2ᵃ Stufe) $\Gamma[2]$ ein ganzes k mit $U^{-2k}L_1 \in \Gamma_\vartheta \cap \Gamma^0[8]$ bestimmen kann, gilt sogar

$$(3.15)\qquad \varphi(\tau)|Z_4 \in \{\Gamma[2],\ -\tfrac{1}{2}k,\ v_\vartheta^k\}.$$

Es sei $k = 8g + r$, $r = 1, 2, 3$, $g \geqq 1$. Nach dem Riemann-Rochschen Satz ist der Rang der Schar der ganzen Spitzenformen in der Klasse (3.15) gleich $2g-1$, also $=1$ für $g=1$. Wenn nun $\varphi(\tau) \in \mathfrak{G}_k^+$ und $g=1$ ist, so erhält man

$$(3.16)\qquad \varphi(\tau)|Z_4 = \gamma_k \varphi(\tau) \quad \text{mit konstantem } \gamma_k \qquad (k = 9, 10, 11)$$

und daher $\beta_{4n} = \gamma_k \beta_n$ $(n \geqq 1)$ oder

$$(3.17)\qquad \beta_1 \beta_{4n} = \beta_4 \beta_n \qquad (n \geqq 1;\ \text{es ist } \beta_1 \neq 0).$$

Dies sind die Relationen (1.12). Für $\varphi(\tau) = f_k(\tau)$ wird $\beta_n = b_k(n)$ und insbesondere

$$(3.18)\qquad b_9(1) = \frac{32}{17}, \quad b_{10}(1) = \frac{32}{5}, \quad b_{11}(1) = \frac{352}{31}.$$

Satz 9. *Der durch* (3.5, 14) *definierte Operator Z_4 führt jede ganze Modulform $\varphi(\tau)$ der Klasse $\mathfrak{K}^k = \{\Gamma_\vartheta, -\frac{1}{2}k, v_\vartheta^k\}$ in eine ganze Modulform $\varphi(\tau)|Z_4$ der Klasse $\{\Gamma[2], -\frac{1}{2}k, v_\vartheta^k\}$ über. Verschwindet φ in allen Spitzen, so auch $\varphi|Z_4$. In diesem Falle gelten für $k = 9, 10, 11$ die Relationen* (3.16, 17) *mit den Werten*

$$\gamma_9 = -8, \quad \gamma_{10} = \gamma_{11} = 16.$$

§ 4. Diskussion der Darstellungen der ρ_m für $m = 4, 5$

4.1. Wir beginnen mit der *Reduktion der $e_k(q)$ auf die $e_k^*(q) = e_k^*$ für $k = 9, 11$ und $q \equiv k \bmod 4$*. Nach [*1*] (22), (38), (39), (40), (46) wird

$$(4.1)\qquad e_9(q) = \begin{cases} \frac{1}{17}\cdot 2^5\cdot 3\cdot 5\cdot 137\, e_9^*, & \text{wenn } q \equiv 1 \bmod 8 \\ 4320\, e_9^*, & \text{wenn } q \equiv 5 \bmod 8 \end{cases}$$

$$e_9^* = \frac{1}{8q}(2q^2 s_2 - s_4) \qquad (q \equiv 1 \bmod 4)$$

und analog

$$e_{11}(q) = \left\{\begin{array}{ll} \frac{1}{31}\, 2^3 \cdot 3 \cdot 7 \cdot 11 \cdot 71\, e_{11}^*, & \text{wenn } q \equiv 3 \bmod 8 \\ 3960\, e_{11}^*, & \text{wenn } q \equiv 7 \bmod 8 \end{array}\right\} \tag{4.2}$$

$$e_{11}^* = \frac{1}{30q}\,(-3s_5 + 10q^2 s_3 - 7q^4 s_1) \qquad (q \equiv 3 \bmod 4).$$

Bereits in [*I*] wurde gezeigt, daß e_9^* ganz ist, wenn $q > 5$. Wir beweisen $e_9^* \equiv \frac{1}{4}(q-1) \bmod 2$, wenn $q > 5$ (vgl. [*I*], S. 373f.). Es ist

$$e_9^* \equiv q\, e_9^* = \tfrac{1}{8}(2q^2 s_2 - s_4) \equiv \tfrac{1}{4} q^2 s_2 \equiv \tfrac{1}{4} s_2 \bmod 2,$$

$$s_2 \equiv 4\,(2/q) \sum_{1 \le j \le \frac{1}{2}(q-1)} (j/q)\, j^2 \bmod 8, \qquad e_9^* \equiv \sum_{1 \le j \le \frac{1}{2}(q-1)} j \equiv \tfrac{1}{4}(q-1) \bmod 2.$$

Wir betrachten jetzt e_{11}^* und zeigen, daß e_{11}^*, wenn $q > 11$, ganz und $\equiv \frac{1}{4}(q+1) \bmod 2$ ist $(q \equiv 3 \bmod 4)$. Es sei also $q > 11$, $q \equiv 3 \bmod 4$. Wir setzen

$$K_{11} = 30q\, e_{11}^* = -3s_5 + 10q^2 s_3 - 7q^4 s_1.$$

Nach bekannten Schlüssen ist $s_5 \equiv 0 \bmod q$, also $K_{11} \equiv 0 \bmod q$. Ferner wird

$$K_{11} \equiv s_5 - s_1 = \sum_{j=1}^{q-1} (j/q)\,(j^5 - j) \equiv 0 \bmod 2$$

$$K_{11} \equiv s_3 - s_1 = \sum_{j=1}^{q-1} (j/q)\,(j^3 - j) \equiv 0 \bmod 3$$

$$K_{11} \equiv 2(s_5 - s_1) = 2\sum_{j=1}^{q-1} (j/q)\,(j^5 - j) \equiv 0 \bmod 5,$$

mithin $K_{11} \equiv 0 \bmod 30q$, $e_{11}^* \equiv 0 \bmod 1$. Zum Beweise der Kongruenz mod 2 beachte man

$$e_{11}^* \equiv 15\, q\, e_{11}^* = \tfrac{1}{2} K_{11} \bmod 2, \qquad K_{11} \equiv s_5 - 2s_3 + s_1 \bmod 4.$$

Nun folgt

$$K_{11} \equiv \sum_{j=1}^{q-1} (j/q)\, j\,(j^2-1)^2 \equiv (2/q) \sum_{j=1}^{\frac{1}{2}(q-1)} (j/q)\, j\,(4j^2-1)^2 \equiv 2 \sum_{j=1}^{\frac{1}{2}(q-1)} j \bmod 4$$

$$e_{11}^* \equiv \tfrac{1}{8}(q^2-1) \equiv \tfrac{1}{4}(q+1) \bmod 2.$$

Zusammen mit $e_9^*(5) = 1$, $e_{11}^*(7) = 16$ ergibt dies

Satz 10. *Für $k = 9, 11$ und alle Primzahlen q, die*

$$q \equiv k \bmod 4, \qquad q > 3, \qquad q \neq 11$$

erfüllen, ist $e_k^* = e_k^*(q)$ ganz und genügt den Kongruenzen

$$e_9^* \equiv \tfrac{1}{4}(q-1) \bmod 2, \qquad e_{11}^* \equiv \tfrac{1}{4}(q+1) \bmod 2.$$

Die Bedeutung dieser e_k^* ergibt sich für die q je einer Restklasse mod 8 aus Satz 6 und (4.1, 2). Danach ist

$$4320\,e_9^* = e_9(q) = a_9(q), \quad \text{wenn} \quad q \equiv 5 \bmod 8, \tag{4.3}$$

bzw.

$$3960\,e_{11}^* = e_{11}(q) = a_{11}(q), \quad \text{wenn} \quad q \equiv 7 \bmod 8, \tag{4.4}$$

die Anzahl der Darstellungen von q als Summe von 9 bzw. 11 Quadraten ganzer Zahlen.

4.2. Wir konstruieren für $m = 4, 5$ kurz den in [*1*] entwickelten Formalismus (vgl. den Kursivtext auf S. 369); der dort auftretende Vorzeichenindex wird weggelassen, wenn er den Wert + hat. − Zunächst ist

$$\beta_1 = \gamma_1, \quad \beta_2 = \tfrac{1}{2}\gamma_1^2 + \gamma_2, \quad \beta_3 = \tfrac{1}{3!}\gamma_1^3 + \gamma_1\gamma_2 + \gamma_3,$$

$$\beta_4 = \tfrac{1}{4!}\gamma_1^4 + \tfrac{1}{2}\gamma_1^2\gamma_2 + \tfrac{1}{2}\gamma_2^2 + \gamma_1\gamma_3 + \gamma_4,$$

$$\beta_5 = \tfrac{1}{5!}\gamma_1^5 + \tfrac{1}{6}\gamma_1^3\gamma_2 + \tfrac{1}{2}\gamma_1\gamma_2^2 + \tfrac{1}{2}\gamma_1^2\gamma_3 + \gamma_2\gamma_3 + \gamma_1\gamma_4 + \gamma_5; \tag{4.5}$$

$$\varrho_4 = -\tfrac{1}{4}\beta_1 + \tfrac{11}{12}\beta_2 - \tfrac{3}{2}\beta_3 + \beta_4,$$

$$\varrho_5 = \tfrac{1}{5}\beta_1 - \tfrac{5}{6}\beta_2 + \tfrac{7}{4}\beta_3 - 2\beta_4 + \beta_5; \tag{4.6}$$

für die γ_ν sind die Werte auf S. 374 einzutragen.

Der *Fall* $m = 4$, $q \equiv 5 \bmod 8$, $q > 5$, den wir jetzt betrachten, vereinfacht sich sehr erheblich dadurch, daß hier γ_3 verschwindet und $\gamma_1 = \frac{1}{4}(q-1)$ ganz rational ist; auf Grund von [*1*], 6. und Satz 10 gilt überdies

$$e_5^* \equiv e_9^* \equiv \gamma_1 \equiv 1 \bmod 2.$$

Nach einiger Rechnung gelangt man zu einer Darstellung

$$\varrho_4 - \binom{\gamma_1}{4} = G_4 + H_4\sqrt{q} \tag{4.7}$$

mit

$$2^9 \cdot 3^2 \cdot 5\,G_4 = -(q^2-1)\big(210(q-5) + 408\big) + 16(q^2-1)^2 + 5 \cdot 24^2 e_5^{*2} q, \tag{4.8}$$

$$3 \cdot 2^6 H_4 = \{5(q^2-1) - 42(q-1) + 88\}\,e_5^* - 8e_9^*. \tag{4.9}$$

Von dem Ausdruck auf der rechten Seite von (4.8) ist mit mäßigem Rechenaufwand zu erkennen, daß er durch $2^9 \cdot 3^2 \cdot 5$ teilbar ist; dies besagt $G_4 \equiv 0$ mod 1. Dagegen erweist sich die rechte Seite von (4.9) als durch 192 teilbar ganz dann und nur dann, wenn

$$e_9^* - e_5^* \equiv 4 \bmod 8 \quad \text{und} \quad e_9^* + e_5^* \equiv 0 \bmod 3 \tag{4.10}$$

zutrifft. Nun ist nach (4.7, 8, 9) $3 \cdot 2^6 \varrho_4$ und nach der Herkunft der ϱ_m auch $V\bar{q}\,\varrho_4$ ganz algebraisch im Zahlkörper $\mathsf{P}(\sqrt{q})$. Mithin ist auch ϱ_4 ganz algebraisch und man erhält zusammenfassend:

Satz 11. *Für* $q \equiv 5 \bmod 8$, $q > 5$ *gilt* $\lambda_{5\pm} \equiv 0 \bmod \sqrt{q}$ *und* (4.10).

4.3. Im *Falle* $m = 5$, $q \equiv 7 \bmod 8$, $q > 7$, den wir abschließend behandeln, ist gegenüber 4.2 zusätzlich (4.5) β_5 zu berücksichtigen. Hier ziehen wir von (4.6) ϱ_5 diejenigen Bestandteile heraus, die von $\beta_1 = \gamma_1$ und von den beiden ersten Gliedern in den Darstellungen der β_g $(g > 1)$ herrühren. So ergibt sich

$$(4.11)\quad \begin{cases} \varrho_5 - \binom{\gamma_1}{5} - \{\gamma_1\}_3 \gamma_2 = \\ \quad = -\gamma_2^2 + \frac{1}{2}\gamma_1\gamma_2^2 + \frac{7}{4}\gamma_3 - 2\gamma_1\gamma_3 + \frac{1}{2}\gamma_1^2\gamma_3 + \gamma_2\gamma_3 - 2\gamma_4 + \gamma_1\gamma_4 + \gamma_5 \end{cases}$$

mit

$$(4.12)\qquad \{\gamma_1\}_3 = \tfrac{1}{12}(2\gamma_1^3 - 12\gamma_1^2 + 21\gamma_1 - 10).$$

In der folgenden Diskussion ist zu beachten, daß nach [*1*], 6. und Satz 10

$$(4.13)\qquad h - 1 \equiv e_7^* \equiv e_{11}^* \equiv 0 \bmod 2, \qquad e_7^* \equiv (q+1)\,h \bmod 3$$

zutrifft. Von den Gliedern auf der rechten Seite von (4.11) ist $-\gamma_2^2$ stets ganz.

Wir beginnen mit $\binom{\gamma_1}{5}$; $\gamma_1 = \frac{q-1}{4} + \frac{h}{2}\sqrt{-q}$ ist eine ganze Zahl in $\mathsf{P}(\sqrt{-q})$. Wegen $\left(\frac{-q}{2}\right) = +1$ zerfällt (2) in zwei verschiedene Primideale in $\mathsf{P}(\sqrt{-q})$ und daher ist $5!\binom{\gamma_1}{5}$ stets durch 8 teilbar. Ebenso ergibt sich

$$5!\binom{\gamma_1}{5} \equiv 0 \bmod 3, \qquad \text{wenn} \qquad q \equiv 2 \bmod 3.$$

Es sei $q \equiv 1 \bmod 3$. Dann ist bereits γ_1 durch 3 teilbar, wenn $h \equiv 0 \bmod 3$. Im Falle $q \equiv 1$, $h \not\equiv 0 \bmod 3$ benutze man

$$\gamma_1 \equiv \gamma_1 - 3 \equiv 4\gamma_1 \equiv -h\sqrt{-q} \equiv (\gamma_1 - 1)^2 \equiv (\gamma_1 - 1)(\gamma_1 - 4) \bmod 3;$$

so findet man $5!\binom{\gamma_1}{5} \equiv 1 + h\sqrt{-q} \bmod 3$. Schließlich gilt stets

$$\gamma_1 \equiv -(q-1) - 2h\sqrt{-q} \bmod 5,$$

also

$$\begin{aligned} 5!\binom{\gamma_1}{5} &\equiv -\prod_{k=0}^{4}(q - 1 + k + 2h\sqrt{-q}) \equiv \\ &\equiv -2h\sqrt{-q}(2h\sqrt{-q}+1)(2h\sqrt{-q}-1)(2h\sqrt{-q}+2)(2h\sqrt{-q}-2) \\ &\equiv -2h\sqrt{-q}(h^2q-1)(h^2q+1) \equiv -2h\sqrt{-q}(q^2-1) \bmod 5. \end{aligned}$$

Zusammenfassend konstatieren wir das folgende Zwischenergebnis:

Hilfssatz 1. Es sei die Primzahl $q \equiv 3 \bmod 4$, $q > 3$, h eine ungerade ganze rationale Zahl, $\gamma_1 = \frac{q-1}{4} + \frac{h}{2}\sqrt{-q}$. Dann gilt $5!\binom{\gamma_1}{5} \equiv 0 \bmod 8$, $5!\binom{\gamma_1}{5} \equiv (1 + h\sqrt{-q})\,h^2(2-q) \bmod 3$, $5!\binom{\gamma_1}{5} \equiv -2h\sqrt{-q}(q^2-1) \bmod 5$.

Wir betrachten den Ausdruck (4.12) $12\{\gamma_1\}_3 \bmod 12$. Wegen $\gamma_1^3 \equiv \gamma_1 \bmod 2$ ist zunächst

$$12\{\gamma_1\}_3 \equiv 2\gamma_1^3 + \gamma_1 - 2 \equiv 3\gamma_1 - 2 \equiv -\gamma_1 - 2 \bmod 4.$$

Für $q \equiv 2 \bmod 3$ wird weiter

$$12\{\gamma_1\}_3 \equiv -\gamma_1^3 - 1 \equiv -\gamma_1 - 1 \equiv 1 + h\sqrt{-q} \bmod 3,$$

für $q \equiv 1 \bmod 3$ dagegen $\gamma_1 \equiv 4\gamma_1 \equiv 2h\sqrt{-q} \bmod 3$, also

$$12\{\gamma_1\}_3 \equiv -\gamma_1^3 - 1 \equiv 2hq\sqrt{-q} - 1 \equiv -1 - h\sqrt{-q} \bmod 3.$$

Insgesamt ergibt sich .

Hilfssatz 2. In den Bezeichnungen von Hilfssatz 1 und (4.12) gilt

$$12\{\gamma_1\}_3 \equiv -\gamma_1 - 2 \bmod 4; \quad 12\{\gamma_1\}_3 \equiv -q(1 + h\sqrt{-q}) \bmod 3.$$

Da ϱ_5 durch Multiplikation einerseits mit $\sqrt{-q}$, andrerseits mit einem Potenzprodukt von 2, 3 und 5 ganz wird, ist ϱ_5 eine ganze Zahl in $\mathsf{P}(\sqrt{-q})$. — Wir trennen nun die Fälle $q \equiv 2$ und $q \equiv 1 \bmod 3$.

4.4 (Fallunterscheidung; Ergebnis). Im *Falle* $q \equiv 2 \bmod 3$ ist γ_3 nach (4.13) ganz; aus der Darstellung (4.11) folgt die Kongruenz

$$(4.14) \qquad \varrho_5 - \binom{\gamma_1}{5} - \{\gamma_1\}_3\gamma_2 \equiv (\tfrac{7}{4} + \tfrac{1}{2}\gamma_1)\gamma_3 + \tfrac{1}{2}\gamma_1\gamma_2 + (\gamma_1 - 2)\gamma_4 + \gamma_5 \bmod 1,$$

die durch $\gamma_1^2 - \gamma_1 \equiv \gamma_2^2 - \gamma_2 \equiv 0 \bmod 2$ vereinfacht wurde. Auf Grund der Formeln [*1*], S. 374, der Hilfssätze und einiger Rechnung ergibt sich zunächst

$$0 = 120\varrho_5 \equiv -2h\frac{e_7^*}{6} + e_{11}^*\sqrt{-q} \equiv 0 \bmod 8,$$

also über (4.13) hinaus

$$(4.15) \qquad e_7^* \equiv e_{11}^* \equiv 0 \bmod 4, \quad e_7^* \equiv e_{11}^* \bmod 8.$$

Nach den Moduln 3 und 5 findet man auf die gleiche Weise mit merklich geringerem Rechenaufwand

$$(4.16) \qquad e_{11}^* \equiv 0 \bmod 3, \quad e_{11}^* \equiv 2(q^2 - 1)h \bmod 5.$$

Im *Falle* $q \equiv 1 \bmod 3$ tritt an die Stelle von (4.14) diejenige Kongruenz, die man aus (4.14) dadurch erhält, daß man den ersten Summanden auf der rechten Seite, also

$$(\tfrac{7}{4} + \tfrac{1}{2}\gamma_1)\gamma_3 \quad \text{durch} \quad (\tfrac{7}{4} - 2\gamma_1 + \tfrac{1}{2}\gamma_1^2 + \gamma_2)\gamma_3$$

ersetzt. Diese Komplikation beruht im wesentlichen darauf, daß im vorliegenden Falle nicht mehr auf die Ganzzahligkeit von γ_3 geschlossen werden kann. Trotzdem verläuft die Diskussion nach dem Modul 8 im wesentlichen wie für $q \equiv 2 \bmod 3$ und ergibt wie dort das Resultat (4.15). Ebenso führt

die sehr einfache Diskussion nach dem Modul 5 direkt auf die zweite Kongruenz (4.16). Nach dem Modul 3 dagegen ist anders zu verfahren. Hier ersetze man e_7^* nach Beseitigung der Nenner gemäß (4.13) überall durch $-h$. Dann ergibt sich $e_{11}^* \equiv h \bmod 3$.

Wir formulieren die Ergebnisse zusammenfassend in (vgl. (1.13))

Satz 12. *Es sei q eine Primzahl $\equiv 7 \bmod 8$, $q > 7$. Zwischen den (von q abhängigen) natürlichen Zahlen h, e_7^*, e_{11}^* bestehen über (4.13) hinaus die folgenden Kongruenzen:*

$$e_7^* \equiv e_{11}^* \equiv 0 \bmod 4, \qquad e_7^* \equiv e_{11}^* \bmod 8,$$

$$e_{11}^* \equiv 2(q+1)\,h \bmod 3, \qquad e_{11}^* \equiv 2(q^2-1)\,h \bmod 5.$$

Überdies gilt

$$\lambda_{5\pm} \equiv 0 \bmod \sqrt{-q}.$$

Literatur

[1] Petersson, Hans: Über eine Zerlegung des Kreisteilungspolynoms von Primzahlordnung. Math. Nachr. **14** (1956). — [2] Siehe Dickson, History of the Theory of Numbers, Bd. II, S. 315. New York: Chelsea Publ. Comp. 1952. — [3] Maass, Hans: Analytische Zahlentheorie II, Vorlesungsausarbeitung. Math. Inst. d. Univ. Heidelberg, W.S. 1950/51. — [4] Pfetzer, Werner: Die Wirkung der Modulsubstitutionen auf mehrfache Thetareihen zu quadratischen Formen ungerader Variablenzahl. Arch. der Math. **4** (1953). — [5] Ebel, Ilse: Analytische Bestimmung der Darstellungsanzahlen natürlicher Zahlen durch spezielle ternäre quadratische Formen mit Kongruenzbedingungen. Math. Z. **64** (1956). — [6] Rau, K. Ananda: On the Representation of a Number as the Sum of an even Number of Squares. J. Madras Univ., B **24**, Nr. 1 (1954). — [7] Wohlfahrt, Klaus: Math. Nachr. **16** (1957).

Münster, II. Mathematisches Institut der Universität

(Eingegangen am 24. 7. 1957)

WALTER, W.
Math. Zeitschr. 71, 308–324 (1959)

Über die Differentialgleichung $u_{xy} = f(x, y, u, u_x, u_y)$*)

Von
WOLFGANG WALTER

I. Eindeutigkeitssätze für die charakteristische Anfangswertaufgabe

§ 1. Einführung

Unter der charakteristischen Anfangswertaufgabe für die Differentialgleichung

$$u_{xy} = f(x, y, u, u_x, u_y) \tag{1}$$

verstehen wir die folgende Aufgabe. Gegeben seien ein abgeschlossenes Rechteck R

$$R: \quad 0 \leqq x \leqq a, \quad 0 \leqq y \leqq b$$

und zwei reellwertige Funktionen $\sigma(x)$, $\tau(y)$ mit $\sigma(0) = \tau(0)$, welche für $0 \leqq x \leqq a$ bzw. $0 \leqq y \leqq b$ stetig differenzierbar sind. Die (reellwertige) Funktion $f(x, y, u, v, w)$ sei für $(x, y) \in R$, $-\infty < u, v, w < \infty$ stetig. Eine Funktion $v(x, y)$ gehöre zur Klasse C^*, wenn v, v_x, v_y und v_{xy} in R existieren und stetig sind. Gesucht ist eine Funktion $u(x, y) \in C^*$, welche Lösung der Differentialgleichung (1) ist und die Anfangswerte

$$u(x, 0) = \sigma(x) \quad \text{für } 0 \leqq x \leqq a, \qquad u(0, y) = \tau(y) \quad \text{für } 0 \leqq y \leqq b$$

annimmt.

Die Frage nach der Existenz einer Lösung wird uns in Teil II beschäftigen. Was die Eindeutigkeit der Aufgabe anbetrifft, so ist nur der klassische Eindeutigkeitssatz bei erfüllter Lipschitz-Bedingung (bezüglich u, v und w; vgl. etwa KAMKE [7, S. 408]) bekannt. Im vorliegenden ersten Teil dieser Arbeit sollen einige allgemeinere Eindeutigkeitskriterien bewiesen werden. Es wird sich insbesondere zeigen, daß die aus der Theorie der gewöhnlichen Differentialgleichungen bekannten Bedingungen von OSGOOD [2] und NAGUMO [4] (vgl. etwa KAMKE [7, S. 141/142]) in entsprechender Übertragung auch hier gelten.

Eine Funktion $u(x, y)$ ist genau dann Lösung unserer Anfangswertaufgabe, wenn sie der Integralgleichung

$$u(x, y) = \chi(x, y) + \int_0^x \int_0^y f(s, t, u, u_x, u_y)\, ds\, dt \tag{2}$$

genügt (KAMKE [7, S. 403 f.]). Dabei ist

$$\chi(x, y) = \sigma(x) + \tau(y) - \sigma(0). \tag{3}$$

*) Diese Arbeit wurde teilweise unterstützt von der United States Air Force (Air Force Office of Scientific Research – Contract AF 49(638)-228). Sie besteht aus insgesamt drei Teilen. Teil II (Existenzsätze für die charakteristische Anfangswertaufgabe) und Teil III (Die nichtcharakteristische Anfangswertaufgabe) werden in Kürze in dieser Zeitschrift erscheinen.

Bevor wir auf diese Gleichung zurückkommen, werden wir im nächsten Paragraphen einige für das folgende wesentliche Hilfsmittel zusammenstellen.

§ 2. Hilfssätze über gewöhnliche Differentialgleichungen

Die folgenden Bezeichnungen werden sich als nützlich erweisen. Die Funktionenklasse G oder genauer $G(0, a)$ umfasse alle reellwertigen Funktionen $g(x, z)$, welche nachstehenden drei Bedingungen genügen:

(α) *$g(x, z)$ ist stetig für* $0<x\leqq a$, $z\geqq 0$;

(β) $g(x, z)\geqq 0$ *und* $g(x, 0)=0$;

(γ) *für jede Lösung $z(x)$ der Differentialgleichung*

$$z'(x)=g(x, z) \tag{4}$$

ist entweder $z\equiv 0$ *oder* $z(x)\geqq \delta x$ $(\delta>0)$.

Gilt außerdem

(δ) *jede Lösung $z(x)$ von (4) ist beschränkt*[1];

(ε) $g(x, z)\leqq g(x, \bar{z})$ *für* $z\leqq \bar{z}$;

so schreiben wir $g\in G'=G'(0, a)$.

Für gewöhnliche Differentialgleichungen besteht der folgende allgemeine Eindeutigkeitssatz von Kamke [7, S. 139][2]:

Die Funktion $f(x, y)$ sei für $0\leqq x\leqq a$, $-\infty<y<+\infty$ *stetig und genüge der Abschätzung*

$$|f(x, y)-f(x, \bar{y})|\leqq g\left(x, |y-\bar{y}|\right) \tag{5}$$

für $0<x\leqq a$, $0<y\leqq b$, *wobei* $g(x, z)\in G(0, a)$ *ist. Dann hat die Anfangswertaufgabe*

$$y'(x)=f(x, y), \qquad y(0)=\eta \tag{6}$$

höchstens eine Lösung.

Den engen Zusammenhang dieses Satzes mit der Frage, wann das Verfahren der sukzessiven Approximation konvergiert, hat zuerst Nagumo [6] erkannt. Die für uns wichtigen Sätze aus der zitierten Arbeit von Nagumo werden nachfolgend in etwas abweichender Formulierung mit Beweisen gegeben.

Satz 1. *Die Funktion $z_0(x)$ sei in* $0\leqq x\leqq a$ *stetig und genüge der Ungleichung*

$$z_0(x)\geqq \int_0^x g\left(t, z_0(t)\right) dt \quad \textit{oder kurz} \quad z_0\geqq K z_0 \quad \textit{mit } g\in G'(0, a) \tag{7}$$

[1]) Da $g(x, z)$ für $x=0$ nicht definiert ist, kann man zunächst von einer Lösung nur im Intervall $0<x\leqq a$ sprechen. Wegen (β) und (δ) läßt sich jede Lösung $z(x)$ nach links bis $x=0$ und nach rechts bis $x=a$ fortsetzen. Wenn im folgenden von einer Lösung $z(x)$ von (4) gesprochen wird, so ist damit immer gemeint, daß $z(x)$ in $[0, a]$ stetig und in $(0, a]$ Lösung der Gleichung (4) ist.

[2]) a. a. O. wird ein noch allgemeinerer, für Systeme gültiger Satz bewiesen; für den Fall *einer* Differentialgleichung findet sich ein ähnlicher Satz bereits bei S. Iyanaga [Japanese J. Math. 6, 253–257 (1928)].

(der Integraloperator K ist im allgemeinen nichtlinear). Ferner sei für jede Lösung $z \not\equiv 0$ von (4) *mit* $z(0)=0$ *(falls eine solche existiert)*

$$z_0(x) < z(x) \quad \text{für } 0 < x < \delta_1 \tag{8}$$

($\delta_1 > 0$ darf von $z(x)$ abhängen). Dann konvergiert die Folge der Funktionen $z_n(x) = K^n z_0(x)$, das ist also die durch

$$z_{n+1}(x) = \int_0^x g\big(t, z_n(t)\big)\,dt \qquad (n = 0, 1, 2, \ldots) \tag{9}$$

definierte Folge, gleichmäßig in $[0, a]$ gegen Null, und dasselbe gilt für jede Folge $\varphi_n(x) = K^n \varphi_0(x)$, wenn φ_0 stetig und $0 \leqq \varphi_0(x) \leqq z_0(x)$ ist.

Der Beweis ist sehr einfach. Die Folge z_n ist monoton fallend; nach (7) ist nämlich $z_0 \geqq z_1$, und aus $z_{n-1} \geqq z_n$ folgt, da $g(x, z)$ in z monoton wachsend ist, sofort $z_n \geqq z_{n+1}$. Die z_n konvergieren also gegen eine Funktion $z(x) \geqq 0$, und aus (9) erhält man durch Grenzübergang, der offenbar unter dem Integralzeichen durchgeführt werden darf, $z = Kz$. Die Funktion $z(x)$ ist also einerseits Lösung von (4) mit $z(0) = 0$, andererseits ist sie wegen der Voraussetzung (8) in einer Umgebung des Nullpunktes kleiner als jede nicht identisch verschwindende Lösung von (4); es ist also $z(x) \equiv 0$. Die Konvergenz ist gleichmäßig, da nach einem bekannten Satz eine monotone Folge stetiger Funktionen mit stetigem Grenzwert immer gleichmäßig konvergiert. Schließlich folgt die letzte Behauptung aus der leicht einzusehenden Beziehung $0 \leqq \varphi_n \leqq z_n$.

Bei den späteren Anwendungen dieses Satzes wird meist φ_0 gegeben sein. Man hat dann zu untersuchen, ob eine Funktion z_0 mit den geforderten Eigenschaften existiert. In dieser Richtung gelten die folgenden Sätze.

SATZ 2. *Die Funktion $\varphi_0(x)$ sei in $[0, a]$ stetig und nichtnegativ. Es gebe eine Lösung $z = \bar{z}(x)$ von* (4) *mit*

$$\varphi_0(x) \leqq \bar{z}(x) \quad \text{für} \quad 0 \leqq x \leqq a, \tag{10}$$

und für jede Lösung $z \not\equiv 0$ von (4) *mit $z(0) = 0$ existiere ein $\delta_1 > 0$ derart, daß*

$$\varphi_0(x) < z(x) \quad \text{für} \quad 0 < x < \delta_1 \tag{8'}$$

ist. Ist dann $g(x, z) \in G'(0, a)$, so konvergiert die Folge von Funktionen

$$\varphi_n(x) = K^n \varphi_0(x)$$

gleichmäßig in $[0, a]$ gegen Null.

BEWEIS. Es genügt, eine Funktion z_0 mit den Eigenschaften (7) und (8) anzugeben. Betrachten wir zuerst den Fall, daß $z \equiv 0$ die einzige Lösung von (4) mit dem Anfangswert $z(0) = 0$ ist. Dann ist (8) von selbst erfüllt, und es genügt, $z_0 = \bar{z}$ zu setzen.

Im anderen Fall gibt es eine Folge von Zahlen c_n mit

$$\bar{z}(a) > c_1 > c_2 > \cdots; \quad c_n \to 0 \quad \text{für} \quad n \to \infty$$

und zugehörigen Funktionen $\zeta_n(x)$ mit den Eigenschaften

$$\zeta_n' = g(x, \zeta_n); \quad \zeta_n(0) = 0; \quad \zeta_n(a) = c_n \qquad (n = 1, 2, 3, \ldots).$$

Die ζ_n seien überdies so gewählt, daß $\zeta_n(x)\leqq\zeta_{n-1}(x)$ ist; wegen der Eigenschaft (ε) von g ist das immer möglich (man nehme etwa die Minimalintegrale). Die durch

$$d_n=\sup\{d\,|\,\varphi_0(x)\leqq\zeta_n(x) \quad \text{für} \quad 0\leqq x\leqq d\}$$

definierte Zahlenfolge ist demnach monoton fallend, und nach (8') ist $d_n>0$ für jedes n. Mit Hilfe dieser Folge wird

$$(11)\qquad z_0(x)=\begin{cases}\bar{z}(x) & \text{für} \quad d_1<x\leqq a\\ \zeta_n(x) & \text{für} \quad d_{n+1}<x\leqq d_n\\ 0 & \text{für} \quad 0\leqq x\leqq \inf d_n\end{cases}\qquad (n=1, 2, \ldots)$$

gesetzt; für diese Funktion ist, wovon man sich leicht überzeugt,

$$z_0\geqq K z_0 \quad \text{und} \quad z_0(x)\geqq\varphi_0(x).$$

Die Funktion z_0 ist im allgemeinen unstetig, hat aber sonst die in Satz 1 geforderten Eigenschaften. Geht man jedoch zu $z_1=Kz_0$ und $\varphi_1=K\varphi_0$ über, so ist z_1 stetig, $\varphi_1\leqq z_1$ und $z_1\geqq Kz_1$, und Satz 1 ist anwendbar mit φ_1, z_1 an Stelle von φ_0, z_0.

Bezüglich der Funktion $g(x, z)$ sind, wie Satz 3 zeigen wird, zwei wesentlich verschiedene Fälle zu unterscheiden. Hierzu einige Bezeichnungen. Wenn $g(x, z)\in G'(0, a)$ den beiden Bedingungen

(ζ) *für jede Lösung* $z(x)$ *von* (4) *ist entweder* $z\equiv 0$ *oder* $z(x)\geqq\delta>0$;

(η) *zu jedem* $C>0$ *gibt es eine Lösung* $z(x)\geqq C$ $(0\leqq x\leqq a)$ *von* (4);

genügt, dann sei $g(x, z)\in G''(0, a)$.

Die Funktionenklasse Φ enthalte alle in $[0, a]$ stetigen, nichtnegativen Funktionen $\varphi(x)$, für welche $\varphi(x)=o(x)$ $(x\to 0)$ ist. Gleichbedeutend ist die Formulierung, daß eine in $[0, a]$ stetige, nichtnegative Funktion $\psi(x)$ existiert, für die $\psi(0)=0$ und $\varphi(x)=x\psi(x)$ ist.

Satz 3. *Für* $g(x, z)\in G'(0, a)$ *und* $\varphi_0(x)\in\Phi$ *und ebenso für* $g(x, z)\in G''(0, a)$ *und jede in* $[0, a]$ *stetige Funktion* $\varphi_0(x)$ *gilt*

$$\varphi_n(x)=K^n\varphi_0(x)\to 0 \quad \textit{für} \quad n\to\infty$$

gleichmäßig in $[0, a]$.

Beweis. Liegt der zweite Fall vor, so ist die Behauptung eine unmittelbare Folge von Satz 2, da die dortige Bedingung (10) nach (η) immer erfüllt werden kann. Im ersten Fall sei $z(x)$ irgendeine nicht identisch verschwindende Lösung von (4). Wegen (γ) und nach der Definition von Φ gibt es ein positives δ_1, so daß $\varphi_0(x)\leqq z(x)$ für $0\leqq x\leqq\delta_1$ ist. Es sei C eine Konstante $\geqq\varphi_0(x)$. Bezeichnen wir das Maximalintegral von (4) durch den Punkt $(\delta_1, C+z(\delta_1))$ mit $\bar{z}(x)$, so hat $\bar{z}(x)$ die in Satz 2 verlangte Eigenschaft $\bar{z}(x)\geqq\varphi_0(x)$.

Betrachten wir hierzu einige Beispiele.

(a) *Lipschitz-Bedingung.* Es sei

$$g(x, z)=K(x)z \quad \text{mit} \quad \int_0^a K(x)\,dx<\infty$$

(natürlich ist $K(x)$ als stetig und nichtnegativ in $(0, a]$ vorausgesetzt). Man sieht leicht, daß $g\in G''$ ist.

(b) Es sei

$$g(x,z) = K(x)\,\omega(z).$$

Über $K(x)$ mögen dieselben Voraussetzungen wie in (a) gelten, $\omega(z)$ sei für $z \geqq 0$ stetig und monoton wachsend, $\omega(0) = 0$, $\omega(z) > 0$ für $z > 0$ und

$$\int_0^1 \omega^{-1}(z)\,dz = \infty.$$

Wieder ist $g(x,z) \in G''$ (*Bedingung von* MONTEL [3], *für* $K(x) = 1$ *von* OSGOOD [2]).

(c) Die Funktion

$$g(x,z) = z/x$$

ist, da alle Lösungen von (4) durch $z = Cx$ gegeben sind, aus der Klasse G' (*Bedingung von* NAGUMO).

(d) Setzen wir

$$g(x,z) = z/x + x\gamma(x, z/x),$$

wobei $\gamma \in G''$ ist, so ist $g \in G'$. Um dies einzusehen, braucht man nur $u(x) = z(x)/x$ zu setzen. Es genügt dann $u(x)$ der Differentialgleichung $u' = \gamma(x,u)$, und nach (δ) und (ζ) ist $u \equiv 0$ oder $0 < \delta \leqq u \leqq C$, also $z \equiv 0$ oder $\delta x \leqq z(x) \leqq Cx$.

Hierin ist insbesondere (c) für $\gamma = 0$ enthalten, und man kann leicht weitere einfache Beispiele erhalten, indem man γ gemäß (a) oder (b) wählt (vgl. dazu NAGUMO [6, S. 159]).

Bei den Sätzen dieses Paragraphen wurde mehr auf einfache Formulierbarkeit als auf größte Allgemeinheit Wert gelegt. So hätte man etwa auch ohne die Voraussetzung (δ) auskommen können. Die Behauptung in Satz 3 müßte dann dahingehend abgeändert werden, daß die φ_n nur in einem eventuell kleineren, von φ_0 abhängenden Intervall $0 \leqq x \leqq a_1$ ($a_1 \leqq a$) konvergieren. Dagegen soll ausdrücklich darauf hingewiesen werden, daß die Monotonievoraussetzung (ε) unentbehrlich ist. Wenn also im folgenden etwa gesagt wird, daß unter der Bedingung von OSGOOD Eindeutigkeit herrscht, so ist das nur cum grano salis richtig; als zusätzliche Voraussetzung wird benötigt, daß $\omega(z)$ monoton wächst (im schwachen Sinne).

§ 3. Die Differentialgleichung $u_{xy} = f(x, y, u)$

In diesem Sonderfall der Gl. (1) ist die Existenz einer Lösung der Integralgleichung (2) gesichert, wenn f stetig und beschränkt ist (ein noch allgemeinerer Existenzsatz findet sich bei ALEXIEWICZ-ORLICZ [9]). Auch die Frage nach der Eindeutigkeit wird hier einfacher und vollständiger als im allgemeinen Fall beantwortet werden können.

SATZ 4. *Die Funktion* $g(x,z) \in G'(0,c)$, $c = ab$, *habe noch die folgende Eigenschaft* (ϑ):

(ϑ) $g(x, xz)$ *ist für jedes feste* $z \geqq 0$ *eine monoton wachsende Funktion von* x.

Dann hat die charakteristische Anfangswertaufgabe für die Differentialgleichung

$$u_{xy} = f(x,y,u) \tag{12}$$

höchstens eine Lösung, wenn die (stetige[3]*) Funktion* $f(x, y, u)$ *für* $0<x\leqq a$, $0<y\leqq b$, $-\infty<u, \bar{u}<+\infty$ *einer Abschätzung*

$$|f(x, y, u)-f(x, y, \bar{u})|\leqq g(xy, \delta u) \qquad (\delta u=|u-\bar{u}|) \tag{13}$$

genügt.

Beweis. Es seien $u(x, y)$ und $\bar{u}(x, y)$ zwei Lösungen. Für die Differenz $\delta u(x, y)=|u(x, y)-\bar{u}(x, y)|$ gilt dann nach (2) und (13) die Abschätzung

$$\delta u(x, y)\leqq\int_0^x\int_0^y |f(s, t, u)-f(s, t, \bar{u})|\,ds\,dt\leqq\int_0^x\int_0^y g\left(st, \delta u(s, t)\right)ds\,dt. \tag{14}$$

Nun ist $u=\bar{u}$ für $x=0$ und für $y=0$ und daher

$$h(x, y)=|f(x, y, u(x, y))-f(x, y, \bar{u}(x, y))|$$

eine in R stetige Funktion, welche für $x=0$ und für $y=0$ verschwindet. Aus (14) folgt

$$\delta u(x, y)\leqq xy\,h(x^*, y^*), \tag{15}$$

wobei $0\leqq x^*\leqq x$, $0\leqq y^*\leqq y$ ist. Die Funktion

$$\psi_0(t)=\sup\{h(x, y)\,|\,xy\leqq t\}$$

ist für $0\leqq t\leqq ab=c$ stetig und monoton wachsend; weiter ist $\psi_0(0)=0$ nach der obigen Bemerkung über $h(x, y)$. Wir haben also die Abschätzung

$$\delta u(x, y)\leqq\varphi_0(xy) \quad \text{mit} \quad \varphi_0(t)=t\,\psi_0(t)\in\Phi.$$

Wegen (ε) und (ϑ) gelten die folgenden Ungleichungen

$$g\left(st, \delta u(s, t)\right)\leqq g\left(st, st\,\psi_0(st)\right)\leqq$$
$$\leqq g\left(xt, xt\,\psi_0(st)\right)\leqq g(xt, xt\,\psi_0(xt)=g\left(xt, \varphi_0(xt)\right).$$

Aus (14) erhält man dann

$$\delta u(x, y)\leqq\int_0^x\int_0^y g\left(xt, \varphi_0(xt)\right)ds\,dt=\int_0^{xy} g\left(r, \varphi_0(r)\right)dr=\varphi_1(xy).$$

Hierbei wurde die in Satz 1 eingeführte Bezeichnung $\varphi_1(t)=K\varphi_0(t)$ übernommen. Mit $g\left(r, \varphi_0(r)\right)$ ist auch der Mittelwert

$$\psi_1(t)=\frac{1}{t}\int_0^t g\left(r, \varphi_0(r)\right)dr$$

monoton wachsend. Mit (A_n) werde die folgende Aussage bezeichnet:

(A_n) Es ist $\delta u(x, y)\leqq\varphi_n(xy)$, wobei $\varphi_n(t)=t\psi_n(t)$ und $\psi_n(t)$ in $[0, c]$ stetig und monoton wachsend ist.

Es wurde — so kann man das bisher Bewiesene zusammenfassen — gezeigt, daß A_0 richtig ist und daß A_1 aus A_0 folgt. In genau derselben Weise wie oben beweist man für jede natürliche Zahl n, daß A_{n+1} aus A_n folgt. Die Aussage A_n ist also für alle n richtig. Da $\varphi_0\in\Phi$ und $g\in G'$ ist, konvergieren die φ_n nach Satz 3 gegen Null, d.h. es ist $\delta u(x, y)=0$, wie behauptet war.

[3]) Es wird immer vorausgesetzt — auch wenn es nicht in jedem Fall formuliert ist —, daß die Funktion f in dem in § 1 angegebenen Bereich stetig ist.

Die in Satz 4 auftretende zusätzliche Bedingung (ϑ) ist weniger einschneidend, als es zunächst den Anschein haben mag. Betrachten wir dazu die Beispiele (a) bis (d) von § 1. Bei (a) wird durch (ϑ) verlangt, daß $xK(x)$ monoton wachsend ist. Das ist der Fall für $K(x)=x^{\alpha}$, $\alpha>-1$ und für $K(x)=[xL(x)]^{-1}$, falls $L(x)$ monoton fällt, insbesondere für alle Funktionen $K(x)$ der „logarithmischen Skala"

$$\frac{1}{x\left(\ln\frac{1}{x}\right)^{1+\varepsilon}},\quad \frac{1}{x\cdot\ln\frac{1}{x}\left(\ln\ln\frac{1}{x}\right)^{1+\varepsilon}},\ldots \qquad (\varepsilon>0).$$

Im Falle (b) gilt (ϑ) sicher dann, wenn $K(x)$ monoton wachsend (insbesondere konstant) ist. Ist aber z.B. $\omega(z)$ so beschaffen, daß $z^{1-\alpha}\omega(z)$ für ein α mit $0<\alpha<1$ monoton wachsend ist — das trifft für alle Funktionen

$$Cz,\quad Cz\ln\frac{1}{z},\quad Cz\ln\frac{1}{z}\ln\ln\frac{1}{z},\ldots$$

zu — so hat man $K(x)$ nur so zu wählen, daß $x^{1-\alpha}K(x)$ monoton wächst.

Bei (c) ist (ϑ) erfüllt, und dasselbe ist auch im Beispiel (d) richtig, wenn $\gamma(x,z)$ so gewählt wird, daß

(ϑ') *$x\,\gamma(x,z)$ monoton wachsend in x ist für jedes feste $z\geqq 0$.*

So ist etwa die Funktion $\gamma(x,z)=K(x)\omega(z)$ zugelassen, falls $xK(x)$ monoton wächst (und außerdem $\omega(z)$ die in (b) genannten Bedingungen erfüllt).

§ 4. Die Differentialgleichung $u_{xy}=f(x,y,u,u_x,u_y)$

Zunächst werde der Eindeutigkeitssatz von NAGUMO betrachtet. Für gewöhnliche Differentialgleichungen ist er, wie PERRON [5] gezeigt hat, ein „bester" Satz in dem Sinne, daß die Eindeutigkeit im allgemeinen verlorengeht, wenn für $g(x,z)$ statt x/z eine Funktion Cx/z mit $C>1$ genommen wird. Es liegt nicht auf der Hand, wie die sachgemäße Verallgemeinerung auf den vorliegenden Fall lautet. Durch Betrachtung einfacher Beispiele wird man auf die folgende Fassung geführt (die auch einen „besten" Satz im obigen Sinne darstellt).

SATZ 5 (*Eindeutigkeitssatz nach* NAGUMO). *Die charakteristische Anfangswertaufgabe für die Gl.* (1) *hat höchstens eine Lösung, wenn die Funktion $f(x,y,u,v,w)$ für $0<x\leqq a$, $0<y\leqq b$, $-\infty<u,v,\ldots,\overline{w}<\infty$ einer Abschätzung*

$$(17)\qquad \begin{cases} |f(x,y,u,v,w)-f(x,y,\bar{u},\bar{v},\bar{w})| \\ \qquad\leqq \alpha\,\dfrac{\delta u}{xy}+\beta\,\dfrac{\delta v}{y}+\gamma\,\dfrac{\delta w}{x} \qquad (\delta u=|u-\bar{u}|,\ldots) \end{cases}$$

mit $\alpha,\beta,\gamma\geqq 0$, $\alpha+\beta+\gamma=1$ genügt.

Sind dagegen α,β,γ drei nichtnegative Zahlen mit $\alpha+\beta+\gamma>1$, so existieren stets beschränkte, stetige Funktionen f, welche der Abschätzung (17) *genügen und für welche die Anfangswertaufgabe mehrere Lösungen hat.*

Wir geben für diesen Satz einen elementaren, von den Betrachtungen des § 2 unabhängigen Beweis. Zunächst überzeugt man sich, daß u, u_x und u_y

auf den Charakteristikenstücken $x=0$ und $y=0$ eindeutig festgelegt ist. So ist etwa $u_x(x, 0)=\sigma'(x)$ und $u_x(0, y)=v(y)$ eine Lösung der gewöhnlichen Differentialgleichung

$$v'(y)=f\big(0, y, \tau(y), v, \tau'(y)\big)=f_1(y, v) \quad \text{mit} \quad v(0)=\sigma'(0).$$

Nach (17) ist aber

$$|f_1(y, v)-f_1(y, \bar{v})| \leqq \beta \frac{\delta v}{y},$$

woraus nach dem Eindeutigkeitssatz von NAGUMO für gewöhnliche Differentialgleichungen in der Tat folgt, daß $v(y)$ eindeutig bestimmt ist. Entsprechend geht der Beweis für u_y.

Sind also $u(x, y)$ und $\bar{u}(x, y)$ zwei Lösungen von (2), so hat die in R stetige Funktion

$$h(x, y)=\big|f\big(x, y, u(x, y), u_x(x, y), u_y(x, y)\big)-f\big(x, y, \bar{u}(x, y), \bar{u}_x(x, y), \bar{u}_y(x, y)\big)\big|$$

die Eigenschaft, daß $h=0$ ist auf den Charakteristikenstücken $x=0$ und $y=0$. Es sei $\psi_0(t)$ eine für $t \geqq 0$ stetige, monoton wachsende und nach oben konvexe Funktion, für welche

$$h(x, y) \leqq \psi_0(x)\, \psi_0(y) \quad \text{und} \quad \psi_0(0)=0 \tag{18}$$

ist. Eine solche Funktion kann man auf folgendem Wege erhalten. Ist

$$\psi(t)=\sup\big\{h(x, y) \,\big|\, 0 \leqq \operatorname{Min}(x, y) \leqq t\big\}$$

für $t \leqq \operatorname{Min}(a, b)$ und $\psi\big(\operatorname{Min}(a, b)\big)=\psi(t)$ für $t>\operatorname{Min}(a, b)$, so that $\bar{\psi}(t)=\sqrt{\psi(t)}$ alle verlangten Eigenschaften bis auf die Konvexität; denn es ist

$$h(x, y) \leqq \psi\big(\operatorname{Min}(x, y)\big)=\bar{\psi}^2\big(\operatorname{Min}(x, y)\big) \leqq \bar{\psi}(x)\, \bar{\psi}(y) \quad \text{und} \quad \psi(0)=\bar{\psi}(0)=0.$$

Es genügt demnach, für $\psi_0(t)$ irgendeine monotone und nach oben konvexe Funktion $\geqq \bar{\psi}(t)$ mit $\psi_0(0)=0$ zu nehmen. Daß eine solche Funktion existiert, ist leicht einzusehen (man betrachte etwa die konvexe Hülle der Ordinatenmenge von $\bar{\psi}$, d.h. der Punktmenge $0 \leqq z \leqq \bar{\psi}(t)$, $t \geqq 0$ in einer (t, z)-Ebene; sie definiert eine Funktion $\psi_0(t)$ mit den gewünschten Eigenschaften). Wegen der Konvexität von $\psi_0(t)$ hat man die Ungleichung

$$\int_0^t \psi_0(s)\, ds \leqq t\, \psi_0(t/2) \quad \text{für} \quad t \geqq 0$$

und allgemeiner, da die Funktion $\psi_n(t)=\psi_0(2^{-n}t)$ (n natürliche Zahl) ebenfalls konvex ist,

$$\int_0^t \psi_n(s)\, ds \leqq t\, \psi_{n+1}(t) \quad \text{für} \quad t \geqq 0. \tag{19}$$

Kehren wir zur Gl. (2) zurück. Mit den Bezeichnungen

$$\delta u(x, y)=|u(x, y)-\bar{u}(x, y)|, \quad \delta u_x(x, y)=|u_x(x, y)-\bar{u}_x(x, y)|,$$
$$\delta u_y(x, y)=|u_y(x, y)-\bar{u}_y(x, y)|$$

(die auch in den späteren Beweisen benutzt werden), ergeben sich aus (2) mit Hilfe von (17) die Ungleichungen

$$(20)\quad \begin{cases} \delta u(x,y) \leqq \int_0^x \int_0^y h(s,t)\,ds\,dt \leqq \int_0^x \int_0^y \left[\alpha \frac{\delta u(s,t)}{st} + \beta \frac{\delta u_x(s,t)}{t} + \gamma \frac{\delta u_y(s,t)}{s}\right] ds\,dt \\ \delta u_x(x,y) \leqq \int_0^y h(x,t)\,dt \leqq \int_0^y \left[\alpha \frac{\delta u(x,t)}{xt} + \beta \frac{\delta u_x(x,t)}{t} + \gamma \frac{\delta u_y(x,t)}{x}\right] dt \\ \delta u_y(x,y) \leqq \int_0^x h(s,y)\,ds \leqq \int_0^x \left[\alpha \frac{\delta u(s,y)}{sy} + \beta \frac{\delta u_x(s,y)}{y} + \gamma \frac{\delta u_y(s,y)}{s}\right] ds. \end{cases}$$

Hieraus und aus (18) erhält man eine erste Abschätzung

$$(21)\quad \begin{cases} \delta u(x,y) \leqq xy\,\psi_0(x)\,\psi_0(y), \quad \delta u_x(x,y) \leqq y\,\psi_0(x)\,\psi_0(y), \\ \delta u_y(x,y) \leqq x\,\psi_0(x)\,\psi_0(y). \end{cases}$$

Nun soll durch Induktion bewiesen werden, daß für $n = 0, 1, 2, \ldots$ die Ungleichungen

$$(22)\quad \begin{cases} \delta u(x,y) \leqq xy \sum_{\nu=0}^{n} \alpha_\nu \psi_\nu(x)\,\psi_{n-\nu}(y) \quad \text{mit} \quad \alpha_\nu \geqq 0,\ \sum_\nu \alpha_\nu = 1 \\ \delta u_x(x,y) \leqq y \sum_{\nu=0}^{n} \beta_\nu \psi_\nu(x)\,\psi_{n-\nu}(y) \quad \text{mit} \quad \beta_\nu \geqq 0,\ \sum_\nu \beta_\nu = 1 \\ \delta u_y(x,y) \leqq x \sum_{\nu=0}^{n} \gamma_\nu \psi_\nu(x)\,\psi_{n-\nu}(y) \quad \text{mit} \quad \gamma_\nu \geqq 0,\ \sum_\nu \gamma_\nu = 1 \end{cases}$$

bestehen. Für $n = 0$ stimmt (22) mit (21) überein. Wir nehmen an, (22) sei für $n = k$ richtig. Dann folgt aus (19), (20) und (22)

$$\delta u(x,y) \leqq \sum_\nu \int_0^x \int_0^y (\alpha\alpha_\nu + \beta\beta_\nu + \gamma\gamma_\nu)\,\psi_\nu(s)\,\psi_{n-\nu}(t)\,ds\,dt \leqq xy \sum_\nu \bar{\alpha}_\nu \psi_\nu(x)\,\psi_{n+1-\nu}(y)$$

$$\delta u_x(x,y) \leqq \sum_\nu \int_0^y (\alpha\alpha_\nu + \beta\beta_\nu + \gamma\gamma_\nu)\,\psi_\nu(x)\,\psi_{n-\nu}(t)\,dt \leqq x \sum_\nu \bar{\alpha}_\nu \psi_\nu(x)\,\psi_{n+1-\nu}(y)$$

$$\delta u_y(x,y) \leqq \sum_\nu \int_0^x (\alpha\alpha_\nu + \beta\beta_\nu + \gamma\gamma_\nu)\,\psi_\nu(s)\,\psi_{n-\nu}(y)\,ds \leqq y \sum_\nu \bar{\alpha}_\nu \psi_{\nu+1}(x)\,\psi_{n-\nu}(y)$$

mit

$$\bar{\alpha}_\nu = \alpha\alpha_\nu + \beta\beta_\nu + \gamma\gamma_\nu, \quad \sum_\nu \bar{\alpha}_\nu = 1,$$

d.h. die Ungleichungen (22) sind auch für $n = k+1$ richtig.

Ist $m = [n/2]$ (größte ganze Zahl $\leqq n/2$) und etwa $a \geqq b$, so gilt für $\nu = 0, 1, \ldots, n$

$$\psi_\nu(x)\,\psi_{n-\nu}(y) \leqq \psi_\nu(a)\,\psi_{n-\nu}(a) \leqq \psi_0(a)\,\psi_m(a),$$

also

$$\delta u(x,y) \leqq xy\,\psi_0(a)\,\psi_m(a) \quad \text{und} \quad \psi_m(a) = \psi_0(2^{-m}a) \to 0 \qquad (n \to \infty).$$

Damit ist gezeigt, daß $u(x,y) = \bar{u}(x,y)$ ist.

Betrachten wir nun für ein noch zu bestimmendes positives ε das folgende Beispiel:

$$u(x, y) = x^{1+\varepsilon} y^{1+\varepsilon} \qquad u_y(x, y) = (1+\varepsilon)\, x^{1+\varepsilon} y^{\varepsilon}$$
$$u_x(x, y) = (1+\varepsilon)\, x^{\varepsilon} y^{1+\varepsilon} \qquad u_{xy}(x, y) = (1+\varepsilon)^2 x^{\varepsilon} y^{\varepsilon}.$$

Es ist

$$u_{xy}(x, y) = \alpha \frac{u}{xy} + \beta \frac{u_x}{y} + \gamma \frac{u_y}{x},$$

wenn α, β, γ drei nichtnegative Zahlen, $\alpha + \beta + \gamma > 1$ und ε die eindeutig bestimmte positive Lösung der Gleichung

$$\alpha + (1+\varepsilon)(\beta + \gamma) = (1+\varepsilon)^2$$

ist. Die rechte Seite dieser Differentialgleichung ist nicht stetig für $(x, y) \in R$ (R sei ein beliebiges Rechteck der in der Einführung angegebenen Art). Doch kann man durch den folgenden einfachen Kunstgriff, der schon von PERRON [5] im Falle der gewöhnlichen Differentialgleichung benützt wurde, zu einer stetigen rechten Seite gelangen. Man setzt

$$f_1(x, y, u) = \begin{cases} 0 & \text{für} \quad u \leqq 0, \\ \dfrac{u}{xy} & \text{für} \quad 0 < u \leqq x^{1+\varepsilon} y^{1+\varepsilon}, \\ x^{\varepsilon} y^{\varepsilon} & \text{für} \quad u \geqq x^{1+\varepsilon} y^{1+\varepsilon}, \end{cases}$$

$$f_2(x, y, v) = \begin{cases} 0 & \text{für} \quad v \leqq 0, \\ \dfrac{v}{y} & \text{für} \quad 0 < v \leqq (1+\varepsilon)\, x^{\varepsilon} y^{1+\varepsilon}, \\ (1+\varepsilon)\, x^{\varepsilon} y^{\varepsilon} & \text{für} \quad v \geqq (1+\varepsilon)\, x^{\varepsilon} y^{1+\varepsilon}, \end{cases}$$

$$f_3(x, y, w) = \begin{cases} 0 & \text{für} \quad w \leqq 0, \\ \dfrac{w}{x} & \text{für} \quad 0 < w \leqq (1+\varepsilon)\, x^{1+\varepsilon} y^{\varepsilon}, \\ (1+\varepsilon)\, x^{\varepsilon} y^{\varepsilon} & \text{für} \quad w \geqq (1+\varepsilon)\, x^{1+\varepsilon} y^{\varepsilon}. \end{cases}$$

Dann ist jede Funktion $C x^{1+\varepsilon} y^{1+\varepsilon}$ mit $0 \leqq C \leqq 1$ eine Lösung der Differentialgleichung

$$u_{xy} = \alpha f_1(x, y, u) + \beta f_2(x, y, u_x) + \gamma f_3(x, y, u_y)$$

aus der Klasse C^* mit identisch verschwindenden Anfangswerten. Die rechte Seite dieser Gleichung ist eine stetige und beschränkte Funktion; sie genügt der Abschätzung (17), aber keiner Abschätzung, bei der α, β, γ durch kleinere Zahlen ersetzt werden. Damit ist Satz 5 vollständig bewiesen.

Nach diesem Ergebnis wird man vermuten, daß ein allgemeinerer Eindeutigkeitssatz von der folgenden Art besteht: Hinreichend für die Eindeutigkeit ist eine Abschätzung

(23) $|f(x, y, u, v, w) - f(x, y, \bar{u}, \bar{v}, \bar{w})| \leqq \alpha\, g(xy, \delta u) + \beta\, g(y, \delta v) + \gamma\, g(x, \delta w)$ ($\alpha, \beta, \gamma \geqq 0$, $\alpha + \beta + \gamma = 1$), wobei $g(x, z) \in G'$ ist. Ein Beweis konnte jedoch nur für den Sonderfall erbracht werden, daß f entweder von u_x oder von u_y unabhängig ist.

SATZ 6. *Die Funktion $f(x, y, u, v)$ sei für $(x, y) \in R$, $-\infty < u, v < +\infty$ stetig und gestatte für $0 < x \leqq a$, $0 < y \leqq b$, $-\infty < u, \ldots, \bar{v} < +\infty$ eine Abschätzung*

$$|f(x, y, u, v) - f(x, y, \bar{u}, \bar{v})| \leqq \alpha\, g(xy, \delta u) + \beta\, g(y, \delta v) \tag{24}$$

mit $\alpha \geqq 0$, $\beta \geqq 0$, $\alpha + \beta = 1$, $\delta u = |u - \bar{u}|$, $\delta v = |v - \bar{v}|$. Dabei genüge die Funktion $g(x, z) \in G'(0, c)$ mit $c = \operatorname{Max}(ab, b)$ noch der Bedingung (ϑ). Dann besitzt die charakteristische Anfangswertaufgabe für die Differentialgleichung

$$u_{xy} = f(x, y, u, u_x) \tag{25}$$

höchstens eine Lösung.

BEWEIS. Wieder sei $\delta u(x, y) = |u(x, y) - \bar{u}(x, y)|$ der Betrag der Differenz zweier Lösungen. Fast wörtlich wie bei Satz 5 zeigt man, daß die Abschätzung (21) auch jetzt gültig ist. Wir benötigen hier indessen nur die offenbar schwächere Aussage, daß eine für $t \geqq 0$ stetige, monoton wachsende Funktion $\psi_0(t)$ mit $\psi_0(0) = 0$ existiert, für welche

$$\delta u(x, y) \leqq xy\,\psi_0(y), \quad \delta u_x(x, y) \leqq y\,\psi_0(y), \quad \text{falls } a \leqq 1, \tag{26}$$

und

$$\delta u(x, y) \leqq xy\,\psi_0(ay), \quad \delta u_x(x, y) \leqq y\,\psi_0(ay), \quad \text{falls } a > 1 \tag{26'}$$

ist.

Der Rest des Beweises wird ähnlich wie bei Satz 4 durchgeführt. Zunächst sei $a < 1$, und wir nehmen an, daß die (für $n = 0$ richtigen) Ungleichungen

$$\left\{\begin{array}{l} \delta u(x, y) \leqq x\,\varphi_n(y), \quad \delta u_x(x, y) \leqq \varphi_n(y) \\ \text{mit} \\ \varphi_n(t) = K^n \varphi_0(t) = t\,\psi_n(t), \quad \varphi_0(t) = t\,\psi_0(t) \end{array}\right. \tag{27}$$

(ψ_n monoton wachsend) für ein festes n bereits bewiesen sind. Dann ist nach (24)

$$\begin{aligned} \delta u_x(x, y) &\leqq \int_0^y |f(x, t, u, u_x) - f(x, t, \bar{u}, \bar{u}_x)|\, dt \\ &\leqq \int_0^y [\alpha\, g(xt, \delta u(x, t)) + \beta\, g(t, \delta u_x(x, t))]\, dt \\ &\leqq \int_0^y [\alpha\, g(xt, x\,\varphi_n(t)) + \beta\, g(t, \varphi_n(t))]\, dt \leqq \int_0^y g(t, \varphi_n(t))\, dt = \varphi_{n+1}(y), \end{aligned}$$

da wegen (ϑ) und $x \leqq 1$

$$g(xt, x\,\varphi_n(t)) = g(xt, xt\,\psi_n(t)) \leqq g(t, t\,\psi_n(t)) = g(t, \varphi_n(t))$$

und $\alpha + \beta = 1$ ist. Setzt man $\varphi_{n+1}(t) = t\psi_{n+1}(t)$, so ergibt sich wörtlich wie bei Satz 4, daß auch $\psi_{n+1}(t)$ monoton wachsend ist. Die zweite Ungleichung in (27) ist also auch für den Index $n + 1$ richtig, und die erste dieser Ungleichungen folgt aus der zweiten durch Integration nach x. Die Abschätzung (27) gilt daher für jedes n, und das ist nur für $\delta u(x, y) = 0$ möglich, da $\varphi_0(t) \in \Phi$ und $g \in G'$ ist und nach Satz 3 die Limesrelation $\varphi_n(t) \to 0$ $(n \to \infty)$ besteht.

Im Falle $a > 1$ lautet die Behauptung entsprechend zu (27)

$$\delta u(x, y) \leqq \frac{x}{a}\,\varphi_n(ay), \quad \delta u_x(x, y) \leqq \frac{1}{a}\,\varphi_n(ay), \tag{27'}$$

und ähnlich wie oben ergibt sich jetzt

$$\delta u_x(x,y) \leqq \int_0^y \left[\alpha\, g\left(x t, x\, \varphi_n(a t)/a\right) + \beta\, g\left(t, \varphi_n(a t)/a\right)\right] dt$$

$$\leqq \int_0^y \left[\alpha\, g\left(a t, \varphi_n(a t)\right) + \beta\, g\left(a t, \varphi_n(a t)\right)\right] dt$$

$$= \frac{1}{a} \int_0^{a y} g\left(s, \varphi_n(s)\right) ds = \frac{1}{a}\, \varphi_{n+1}(a y)\,.$$

Hieraus folgt genau wie im ersten Fall $\delta u(x, y) = 0$.

Zusatz. Ein entsprechender Eindeutigkeitssatz gilt natürlich auch für die Differentialgleichung

$$u_{xy} = f(x, y, u, u_y)\,, \tag{25'}$$

falls für f die Abschätzung

$$|f(x, y, u, w) - f(x, y, \bar{u}, \bar{w})| \leqq \alpha\, g(x y, \delta u) + \gamma\, g(x, \delta w) \qquad (\alpha + \gamma = 1) \tag{24'}$$

besteht.

Satz 7. *Die Funktion $g(x, z)$ sei so beschaffen, daß $g_1(x, z) = (2+a)\, g(x, z) \in G'(0, a+b)$ ist. Dann ist die Lösung der charakteristischen Anfangswertaufgabe für die Differentialgleichung* (1) *eindeutig bestimmt, wenn $f(x, y, u, v, w)$ der Ungleichung*

$$\begin{cases} |f(x, y, u, v, w) - f(x, y, \bar{u}, \bar{v}, \bar{w})| \\ \qquad\qquad \leqq g(x+y, \delta u + \delta v + \delta w) \qquad (\delta u = |u - \bar{u}|, \ldots) \end{cases} \tag{28}$$

(für alle $(x, y) \in R$ und alle reellen $u, \ldots, \bar{w}$) genügt.

Beweis. Es seien $u(x, y)$ und $\bar{u}(x, y)$ zwei Lösungen und

$$D(x, y) = \delta u(x, y) + \delta u_x(x, y) + \delta u_y(x, y)\,.$$

Ähnlich wie bei Satz 5 überlegt man sich, daß u_x und u_y auf den Charakteristikenstücken $x = 0$ und $y = 0$ durch die Bedingung (28) eindeutig festgelegt sind und daß $D(x, y)$ einer Abschätzung

$$D(x, y) \leqq \varphi_0(x+y) \quad \text{mit} \quad \varphi_0(t) \in \Phi \tag{29}$$

genügt. Es sei K_1 der (gemäß (7)) zur Funktion $g_1(x, z)$ gehörige Integraloperator und $\varphi_n(t) = K_1^n \varphi_0(t)$. Es soll durch vollständige Induktion gezeigt werden, daß

$$D(x, y) \leqq \varphi_n(x+y) \qquad \text{für } n = 0, 1, 2, \ldots \tag{30}$$

ist. Für $n = 0$ ist die Behauptung nach (29) richtig; ist sie für ein festes n schon bewiesen, so folgt aus

$$\delta u(x, y) \leqq \int_0^{x y}\int_0^{x y} g\left(s+t, D(s, t)\right) ds\, dt \leqq \int_0^{x y}\int_0^{x y} g\left(s+t, \varphi_n(s+t)\right) ds\, dt \leqq \int_0^{x+y} g\left(\tau, \varphi_n(\tau)\right) d\tau$$

$$\delta u_x(x, y) \leqq \int_0^{y} g\left(x+t, D(x, t)\right) dt \leqq \int_0^{y} g\left(x+t, \varphi_n(x+t)\right) dt \leqq \int_0^{x+y} g\left(\tau, \varphi_n(\tau)\right) d\tau$$

$$\delta u_y(x, y) \leqq \int_0^{x} g\left(s+y, D(s, y)\right) ds \leqq \int_0^{x} g\left(s+y, \varphi_n(s+y)\right) ds \leqq \int_0^{x+y} g\left(\tau, \varphi_n(\tau)\right) d\tau$$

die Ungleichung

$$D(x,y) \leqq (2+a)\int_0^{x+y} g\big(\tau, \varphi_n(\tau)\big)\,d\tau = K_1\,\varphi_n(x+y) = \varphi_{n+1}(x+y).$$

Wegen $\varphi_n(t) \to 0$ $(n \to \infty)$ folgt aus (30) die Behauptung $D(x,y) = 0$.

Der Satz bleibt gültig, wenn in der Voraussetzung die Konstante $2+a$ durch $2+b$ ersetzt wird (bei der Abschätzung von δu im Beweis darf der Faktor x vor dem letzten Integral durch y ersetzt werden). Außerdem sei bemerkt, daß die Bedingung (ϑ) zum Beweis nicht benötigt wurde.

Wir kommen noch einmal auf die im Zusammenhang mit (23) geäußerte Vermutung zurück. Satz 7 bezieht sich auch auf die allgemeine Gl. (1), doch ist die Voraussetzung (28) schärfer als (23). Wählt man für $g(x,z)$ in Satz 7 etwa die Funktion $\frac{z}{(2+a)\,x}$, so erhält man einen „Eindeutigkeitssatz nach NAGUMO", dessen Aussage aber schwächer als die von Satz 5 ist. Dagegen ergeben sich interessante Spezialfälle, wenn man für g Funktionen der in § 2, Beispiele (a) und (b) angegebenen Art nimmt. Das Beispiel (a) führt auf den seit langem bekannten Eindeutigkeitssatz bei erfüllter Lipschitz-Bedingung (mit der Verallgemeinerung, daß $K(x)$ keine Konstante zu sein braucht). Mit Beispiel (b) erhalten wir eine

Übertragung des Eindeutigkeitssatzes von OSGOOD: *Die charakteristische Anfangswertaufgabe für die Gl.* (1) *hat höchstens eine Lösung, wenn* $f(x,y,u,v,w)$ *der Ungleichung*

$$(31)\qquad \begin{cases} |f(x,y,u,v,w) - f(x,y,\bar u,\bar v,\bar w)| \\ \qquad\qquad \leqq \omega(\delta u + \delta v + \delta w) \qquad\qquad (\delta u = |u-\bar u|, \ldots) \end{cases}$$

genügt, wobei $\omega(z)$ *für* $z \geqq 0$ *stetig und monoton wachsend,* $\omega(0) = 0$ *und* $\omega(z) > 0$ *für* $z > 0$ *und das Integral*

$$\int_0^1 \frac{dz}{\omega(z)}$$

divergent ist.

§ 5. Verallgemeinerungen und Bemerkungen

Es seien $u = (u_1, \ldots, u_m)$ und entsprechend v, w, f Vektoren eines m-dimensionalen Vektorraumes R_m; der Vektor $f = f(x,y,u,v,w)$ hänge in dem Bereich

$$(x,y) \in R, \qquad -\infty < u_i, v_i, w_i < +\infty \qquad (i = 1, \ldots, m)$$

(R sei ein Rechteck von der in § 1 angegebenen Art) stetig von allen $(3m+2)$ Veränderlichen ab. Mit diesen Bezeichnungen ist (1), oder ausführlicher

$$(1')\qquad \partial^2 u_i(x,y)/\partial x\,\partial y = f_i\big(x,y,u_1(x,y), \ldots, \partial u_m(x,y)/\partial y\big) \qquad (i = 1, \ldots, m)$$

ein hyperbolisches System von Differentialgleichungen, und die in § 1 formulierte Aufgabe ist eine charakteristische Anfangswertaufgabe für dieses System, wenn weiter angenommen wird, daß $\sigma(x)$ und $\tau(y)$ jetzt m-dimensionale Vektorfunktionen sind.

Will man die Eindeutigkeitssätze aus § 3 und § 4 auf Systeme übertragen, so hat man zunächst zu überlegen, was unter $|u|$ verstanden werden soll. Bei Systemen von gewöhnlichen Differentialgleichungen

$$y' = f(x, y) \tag{32}$$

(hierbei sind also y und f Vektoren aus R_m, x ist eine reelle Veränderliche) wird in der Literatur häufig die Normierung

$$|u| = |u_1| + \cdots + |u_m| \tag{33}$$

verwendet (PERRON [5], NAGUMO [6], KAMKE [7], um nur einige zu nennen). So ist etwa der in § 2 zitierte Eindeutigkeitssatz für gewöhnliche Differentialgleichungen auch für Systeme richtig, wenn in der Gl. (5) die Absolutstriche gemäß (33) interpretiert werden (KAMKE [7, S. 139]). Es ist jedoch bei dem zu beweisenden Eindeutigkeitssatz für das System (1) nicht nötig, sich auf eine bestimmte Norm festzulegen.

Dementsprechend sei der Raum R_m in irgendeiner Weise normiert, d.h. $\|u\|$ sei für jedes $u \in R$ erklärt und habe die üblichen Normeneigenschaften

$$0 \leqq \|u + v\| \leqq \|u\| + \|v\|, \qquad \|\alpha u\| = |\alpha| \|u\| \quad \text{für reelles } \alpha,$$
$$\|u\| = 0 \text{ genau dann, wenn } u = 0.$$

Es gilt dann der

SATZ 8. *Die Eindeutigkeitssätze von § 3 und § 4 gelten auch, wenn (1) ein System von Differentialgleichungen ist und wenn an den entsprechenden Stellen — es sind dies die Ungleichungen* (13), (17), (24), (28) —

$$|f(x, y, u, \ldots) - f(x, y, \bar{u}, \ldots)| \quad \textit{durch} \quad \|f(x, y, u, \ldots) - f(x, y, \bar{u}, \ldots)\|$$

sowie

$$\delta u = |u - \bar{u}|, \ldots \quad \textit{durch} \quad \delta u = \|u - \bar{u}\|, \ldots$$

ersetzt wird.

BEWEIS. Ist e_i der Vektor, der an der i-ten Stelle die Komponente 1 und an allen anderen Stellen die Komponente 0 hat ($i = 1, \ldots, m$), so folgt aus der Zerlegung $u = u_1 e_1 + \cdots + u_m e_m$ die Ungleichung

$$\begin{cases} \|u\| \leqq |u_1| \|e_1\| + \cdots + |u_m| \|e_m\| \\ \qquad\quad \leqq K(|u_1| + \cdots + |u_m|) \qquad \text{mit} \quad K = \max_i \|e_i\|. \end{cases} \tag{34}$$

Es seien $u(x, y)$ und $\bar{u}(x, y)$ zwei Lösungen der charakteristischen Anfangswertaufgabe, und wir nehmen für den Augenblick an (was noch zu beweisen sein wird), daß sie samt ihren Ableitungen erster Ordnung auf den Charakteristikenstücken $x = 0$ und $y = 0$ übereinstimmen. Ähnlich wie beim Beweis zu Satz 4 und 5 führen wir eine Funktion $h(x, y)$ ein:

$$h(x, y) = \|f(x, y, u(x, y), \ldots) - f(x, y, \bar{u}(x, y), \ldots)\|$$
$$\leqq K \sum_{i=1}^{n} |f_i(x, y, u, \ldots) - f_i(x, y, \bar{u}, \ldots)|.$$

Die Funktion $h(x, y)$ ist auch jetzt stetig in R und $=0$ für $x=0$ und für $y=0$. Da ferner für jede integrierbare Vektorfunktion $F(s)$

$$\left\| \int_\alpha^\beta F(s)\,ds \right\| \leqq \int_\alpha^\beta \|F(s)\|\,ds \tag{35}$$

ist (entsprechendes gilt für mehrfache Integrale), haben wir

$$\begin{aligned} \delta u(x,y) &= \|u(x,y) - \bar{u}(x,y)\| \\ &\leqq \int_0^x \int_0^y \|f(s,t,u,\ldots) - f(s,t,\bar{u},\ldots)\|\,ds\,dt \leqq \int_0^x \int_0^y h(s,t)\,ds\,dt, \end{aligned}$$

woraus genau wie früher eine erste Abschätzung

$$\delta u(x,y) \leqq \varphi_0(xy) \quad \text{(Satz 4)} \qquad \text{bzw.} \qquad \delta u(x,y) \leqq xy\,\psi_0(x)\,\psi_0(y) \quad \text{(Satz 5)}$$

mit den dort verlangten Eigenschaften existiert. Im übrigen können die früheren Beweise ohne Änderung übernommen werden.

Satz 4 ist damit vollständig bewiesen, bei den Sätzen des § 4 fehlt noch der Nachweis, daß $u_x(0, y) = v(y)$ und $u_x(x, 0) = w(x)$ eindeutig bestimmt sind. Wie beim Beweis von Satz 5 schon festgestellt wurde, genügt die Funktion $v(y)$ den Gleichungen

$$v'(y) = f\big(0, y, \tau(y), v, \tau'(y)\big) = f_1(y, v) \qquad \text{und} \qquad v(0) = \sigma'(0) \tag{36}$$

(ähnliches gilt für $w(x)$). Wir benötigen also den folgenden

Eindeutigkeitssatz für Systeme gewöhnlicher Differentialgleichungen: Es seien $y = (y_1, \ldots, y_m)$ und $f = (f_1, \ldots, f_m)$ Vektoren eines in beliebiger Weise normierten R_m (die Norm sei mit $\|y\|$ bezeichnet). Die Funktion $f = f(x, y)$ sei für

$$0 \leqq x \leqq a, \qquad -\infty < y_i < +\infty \qquad (i = 1, \ldots, m)$$

stetig und genüge der Abschätzung

$$\|f(x,y) - f(x,\bar{y})\| \leqq g(x, \delta y) \qquad (\delta y = \|y - \bar{y}\|),$$

wobei $g(x, z)$ eine Funktion aus der Klasse $G'(0, a)$ ist. Dann besitzt die Anfangswertaufgabe

$$y' = f(x, y), \qquad y(0) = \eta \tag{37}$$

bei vorgegebenem $\eta = (\eta_1, \ldots, \eta_m)$ nicht mehr als eine Lösung.

Da sich ein solcher Satz in der Literatur nicht findet, sei der Beweis kurz skizziert. Es seien $y(x), \bar{y}(x)$ zwei für $0 \leqq x \leqq a_1$ $(a_1 \leqq a)$ existierende Lösungen von (37); es soll gezeigt werden, daß $\delta y(x) = \|y(x) - \bar{y}(x)\| = 0$ ist für $0 \leqq x \leqq a_1$. Aus den Ungleichungen

$$\delta y(x) = \left\| \int_0^x [f(t,y) - f(t,\bar{y})]\,dt \right\| \leqq \int_0^x \|f(t,y) - f(t,\bar{y})\|\,dt \leqq \int_0^x g\big(t, \delta y(t)\big)\,dt$$

leitet man leicht ab, daß erstens

$$\delta y(x) \leqq \varphi_0(x) \qquad \text{für} \qquad 0 \leqq x \leqq a_1 \qquad \text{mit} \qquad \varphi_0(x) \in \Phi$$

ist und daß zweitens, wenn wie früher $\varphi_n(x)=K^n\varphi_0(x)$ gesetzt wird, aus $\delta y(x)\leqq\varphi_n(x)$ die Ungleichung $\delta y(x)\leqq\varphi_{n+1}(x)$ folgt. Nach Satz 3 folgt daraus $\delta y(x)=0$ oder, was dasselbe ist, $y(x)=\bar{y}(x)$ für $0\leqq x\leqq a_1$.

Damit ist auch der Beweis von Satz 8 abgeschlossen.

Hierzu ein einfaches Beispiel. Es bedeute $[z/x]$ eine für $x\geqq 0$ stetige Funktion, welche für negative z verschwindet und für kleine positive z mit z/x übereinstimmt, etwa (wenn $\varphi(x)$ stetig und $\in\Phi$ ist)

$$[z/x]=0 \quad \text{bzw.} \quad z/x \quad \text{bzw.} \quad \varphi(x)/x$$

für

$$z\leqq 0 \quad \text{bzw.} \quad 0<z\leqq\varphi(x) \quad \text{bzw.} \quad z\geqq\varphi(x).$$

Es sei das gewöhnliche System

$$y_1'=\alpha[y_2/x], \qquad y_2'=\beta[y_1/x]$$

vorgelegt, und es soll untersucht werden, für welche Zahlen α, β die Lösung $y_1(x)=y_2(x)=0$ die einzige Lösung mit dem Anfangswert $y_1(0)=y_2(0)=0$ ist. Nach dem Satz von NAGUMO in der üblichen Form (d.h. mit der Norm $\|y\|=|y_1|+|y_2|$) ergibt sich als hinreichende Bedingung für die Einzigkeit

$$|\alpha\,\delta y_2/x|+|\beta\,\delta y_1/x|\leqq(\delta y_1+\delta y_2)/x, \qquad \text{also} \qquad |\alpha|\leqq 1,\ |\beta|\leqq 1$$

($\delta y_i=|y_i-\bar{y}_i|$). Für die Norm $\|y\|=|y_1|+c\,|y_2|$ ($c>0$) lautet die Bedingung dagegen

$$|\alpha\,\delta y_2/x|+|\beta c\,\delta y_1/x|\leqq(\delta y_1+c\,\delta y_2)/x, \qquad \text{also} \qquad |\alpha|\leqq c,\ |\beta|\leqq 1/c.$$

Analoges gilt für das hyperbolische System

$$\frac{\partial^2 u_1}{\partial x\,\partial y}=\alpha\left[\frac{u_2}{xy}\right], \quad \frac{\partial^2 u_2}{\partial x\,\partial y}=\beta\left[\frac{u_1}{xy}\right]$$

Abschließend sollen noch zwei allgemeine, auf das Eindeutigkeitsproblem bezügliche Fragen aufgeworfen werden. Bei gewöhnlichen Differentialgleichungen weiß man, daß, grob gesprochen, die Eindeutigkeit „im großen" aus der Eindeutigkeit „im kleinen" folgt und daß es deshalb hinreichend ist, wenn Abschätzungen der Form (5) lokal gelten. Weiter ist bekannt, daß es genügt, statt einer Abschätzung des Betrages wie in (5) eine „einseitige" Abschätzung

$$f(x, y)-f(x, \bar{y})\leqq g(x, y-\bar{y}) \qquad \text{für} \qquad y>\bar{y}$$

zu fordern. Es ist ein im wesentlichen ungelöstes Problem, ob die hier gegebenen Eindeutigkeitssätze ebenfalls in dieser Richtung verallgemeinert werden können.

Was die erste Frage anbetrifft, so sieht man unschwer, daß die hier bewiesenen Eindeutigkeitssätze richtig bleiben, wenn bezüglich *einer* Koordinate nur eine lokale Abschätzung gefordert wird. Am Beispiel des Satzes 5 sei dies genauer formuliert. Statt (17) möge gelten:

Zu jedem x_0 mit $0 \leqq x_0 < a$ gibt es ein $\delta > 0$, so daß

$$(17') \quad |f(x,y,u,v,w) - f(x,y,\bar{u},\bar{v},\bar{w})| \leqq \alpha \frac{\delta u}{(x-x_0)\,y} + \beta \frac{\delta v}{y} + \gamma \frac{\delta w}{x-x_0}$$

für $x_0 < x < x_0+\delta$, $0 < y \leqq b$, $-\infty < u, v, \ldots, \bar{w} < +\infty$ ist.
Dann hat die charakteristische Anfangswertaufgabe höchstens eine Lösung.

Ist nämlich $\delta u(x, y)$ der Betrag der Differenz zweier Lösungen und

$$x_0 = \sup\{\xi \mid \delta u(x,y) = 0 \quad \text{für} \quad 0 \leqq x \leqq \xi,\ 0 \leqq y \leqq b\},$$

so erhält man aus der Annahme $x_0 < a$ leicht einen Widerspruch, wenn man den Beweis von Satz 5 mit dem Rechteck

$$R'\colon\ x_0 \leqq x \leqq x_0 + \delta/2, \quad 0 \leqq y \leqq b$$

durchführt.

Literatur

[1] PICARD, E.: Sur les méthodes d'approximations successives dans la théorie des équations différentielles. (Note I in Bd. 4 von G. DARBOUX, Leçons sur la théorie générale des surfaces, Paris 1896, S. 353–367.) — [2] OSGOOD, W. F.: Beweis der Existenz einer Lösung der Differentialgleichung $dy/dx = f(x, y)$ ohne Hinzunahme der Cauchy-Lipschitzschen Bedingung. Monatshefte Math. u. Physik **9**, 331–345 (1898). — [3] MONTEL, P.: Sur l'intégrale supérieure et l'intégrale inférieure d'une équation différentielle. Bull. des sciences math. **50**, 205–217 (1926), bes. 215. — [4] NAGUMO, M.: Eine hinreichende Bedingung für die Unität der Lösung von Differentialgleichungen erster Ordnung. Japanese J. Math. **3**, 107–112 (1926). — [5] PERRON, O.: Eine hinreichende Bedingung für die Unität der Lösung von Differentialgleichungen erster Ordnung. Math. Z. **28**, 216–219 (1928). — [6] NAGUMO, M.: Über das Verfahren der sukzessiven Approximationen zur Integration gewöhnlicher Differentialgleichung und die Eindeutigkeit ihrer Integrale. Japanese J. Math. **7**, 143–160 (1930). — [7] KAMKE, E.: Differentialgleichungen reller Funktionen. Leipzig: Akademische Verlagsgesellschaft 1930. — [8] HARTMAN, P., and A. WINTNER: On hyperbolic partial differential equations. Amer. J. Math. **74**, 834–864 (1952). — [9] ALEXIEWICZ, A., and W. ORLICZ: Some remarks on the existence and uniqueness of solutions of the hyperbolic equation $\partial^2 z/\partial x\,\partial y = f(x, y, z, \partial z/\partial x, \partial z/\partial y)$. Studia Mathematica **15**, 201–215 (1956). — [10] DIAZ, J. B.: On an analoge of the Euler-Cauchy polygon method for the numerical solution of $u_{xy} = f(x, y, u, u_x, u_y)$. Arch. Rational Mechanics and Analysis **1**, 357–390 (1958).

Institute for Fluid Dynamics and Applied Mathematics
University of Maryland, College Park/Md. (USA)
und
Institut für Angewandte Mathematik der Technischen Hochschule Karlsruhe

(Eingegangen am 12. November 1958)

Baer, R.
Math. Zeitschr. 71, 325–334 (1959)

Kriterien für die Abgeschlossenheit endlicher Gruppen

Von
Reinhold Baer

Ist $\mathfrak{r}$ eine Menge von Primzahlen, so ist das Element g in der endlichen Gruppe G ein $\mathfrak{r}$-Element, wenn alle Primteiler seiner Ordnung $o(g)$ in $\mathfrak{r}$ liegen. Entsprechend heißt die Gruppe G eine $\mathfrak{r}$-Gruppe, wenn alle Primteiler ihrer Ordnung $o(G)$ zu $\mathfrak{r}$ gehören; und dies ist äquivalent mit der Eigenschaft, daß die Gruppe G nur aus $\mathfrak{r}$-Elementen besteht. Allgemeiner wollen wir unter einer *$\mathfrak{r}$-abgeschlossenen Gruppe eine Gruppe G derart verstehen, daß Produkte von $\mathfrak{r}$-Elementen aus G wieder $\mathfrak{r}$-Elemente sind*; und dies ist äquivalent mit der Forderung, daß die Gesamtheit $G_\mathfrak{r}$ der $\mathfrak{r}$-Elemente aus G eine charakteristische $\mathfrak{r}$-Untergruppe ist [deren Index $[G:G_\mathfrak{r}]$ dann natürlich zu jeder Primzahl aus $\mathfrak{r}$ teilerfremd ist]. Aus der $\mathfrak{r}$-Abgeschlossenheit der Gruppe folgt also insbesondere, daß die Anzahl der $\mathfrak{r}$-Elemente in G genau das kleinste gemeinschaftliche Vielfache $o(G)_\mathfrak{r}$ aller nur durch Primzahlen aus $\mathfrak{r}$ teilbaren Teiler der Ordnung $o(G)$ ist; und Frobenius hat die bisher weder bewiesene noch widerlegte Vermutung ausgesprochen, daß diese Bedingung auch hinreichend sei. Nun ist diese Frobeniussche Anzahlbedingung sicherlich dann hinreichend, wenn die betrachtete Gruppe auflösbar ist, wie sich sofort aus den Hallschen Sätzen ergibt; oder wenn $\mathfrak{r}$ nur aus einer Primzahl besteht, wie aus den Sylowschen Sätzen folgt. Dies legt es nahe, nach ausreichenden Zusatzbedingungen zu suchen; und dies wollen wir im folgenden tun.

Untergruppen $\mathfrak{r}$-abgeschlossener Gruppen sind natürlich auch $\mathfrak{r}$-abgeschlossen, so daß auch die Untergruppen $\mathfrak{r}$-abgeschlossener Gruppen die Frobeniussche Anzahleigenschaft haben. Wir wollen deshalb die Gruppe G *schwach $\mathfrak{r}$-abgeschlossen nennen, wenn in jeder Untergruppe U von G die Anzahl der $\mathfrak{r}$-Elemente genau $o(U)_\mathfrak{r}$ ist.* Es ist klar, daß jede $\mathfrak{r}$-abgeschlossene Gruppe auch schwach $\mathfrak{r}$-abgeschlossen ist. Ob die Umkehrung hiervon gilt, wissen wir nicht; ebenso ist uns unbekannt, ob die Gruppe G schwach $\mathfrak{r}$-abgeschlossen ist, wenn die Anzahl der $\mathfrak{r}$-Elemente in G genau $o(G)_\mathfrak{r}$ ist. Doch sollen hier wenigstens die folgenden Kriterien bewiesen werden.

Satz A. *Dann und nur dann ist G eine $\mathfrak{r}$-abgeschlossene Gruppe mit nilpotentem $G/G_\mathfrak{r}$, wenn G schwach $\mathfrak{r}$-abgeschlossen ist und alle $\mathfrak{r}'$-Untergruppen von G nilpotent sind.*

Hierbei sei — wie auch stets im folgenden — unter $\mathfrak{r}'$ die zu $\mathfrak{r}$ komplementäre Primzahlmenge verstanden.

Satz B. *Dann und nur dann ist G eine $\mathfrak{r}$-abgeschlossene Gruppe mit nilpotentem $G_\mathfrak{r}$, wenn G schwach $\mathfrak{r}$-abgeschlossen ist und alle $\mathfrak{r}$-Untergruppen von G nilpotent sind.*

SATZ C. *Dann und nur dann ist G eine $\mathfrak{r}$-abgeschlossene Gruppe mit auflösbarem $G/G_{\mathfrak{r}}$, wenn G schwach $\mathfrak{r}$-abgeschlossen und $o(E)$ für jede einfache Faktorgruppe E einer Untergruppe von G durch höchstens eine Primzahl aus $\mathfrak{r}'$ teilbar ist.*

SATZ D. *Dann und nur dann ist G eine $\mathfrak{r}$-abgeschlossene Gruppe mit auflösbarem $G_{\mathfrak{r}}$, wenn G schwach $\mathfrak{r}$-abgeschlossen und $o(E)$ für jede einfache Faktorgruppe E einer Untergruppe von G durch höchstens eine Primzahl aus $\mathfrak{r}$ teilbar ist.*

Es sei noch auf die Charakterlosigkeit unserer Überlegungen hingewiesen. Denn sowohl bei den von uns geführten Beweisen wie auch bei den Beweisen der von uns benutzten Sätze wird von der Theorie der Gruppencharaktere u.ä. kein Gebrauch gemacht.

Bezeichnungen

$o(G)$ = Ordnung von G.

$o(g)$ = Ordnung des Gruppenelements g.

$\mathfrak{N}U$ = Normalisator der Untergruppe U [in G].

$\mathfrak{Z}U$ = Zentralisator der Untergruppe U [in G].

$\mathfrak{r}$-Element = Element, dessen Ordnung nur durch Primzahlen aus der Menge $\mathfrak{r}$ teilbar ist.

$\mathfrak{r}$-Gruppe = Gruppe, deren Ordnung nur durch Primzahlen aus der Menge $\mathfrak{r}$ teilbar ist.

$G_{\mathfrak{r}}$ = Menge der $\mathfrak{r}$-Elemente in der Gruppe G.

$n_{\mathfrak{r}}$ = k.g.V. aller nur durch Primzahlen aus $\mathfrak{r}$ teilbaren Teiler von n.

$\mathfrak{r}'$ = zur Primzahlmenge $\mathfrak{r}$ komplementäre Primzahlmenge.

1. Wir beginnen unsere Überlegungen mit der Herleitung einiger allgemeiner Eigenschaften der schwachen $\mathfrak{r}$-Abgeschlossenheit.

LEMMA 1. *Untergruppen und Faktorgruppen schwach $\mathfrak{r}$-abgeschlossener Gruppen sind schwach $\mathfrak{r}$-abgeschlossen.*

BEWEIS. Daß die Untergruppen schwach $\mathfrak{r}$-abgeschlossener Gruppen ebenfalls schwach $\mathfrak{r}$-abgeschlossen sind, ist ohne weiteres aus der Definition zu entnehmen.

Sei weiter K ein Normalteiler der schwach $\mathfrak{r}$-abgeschlossenen Gruppe G. Erzeugen dann die Elemente X und Y aus G/K die gleiche zyklische $\mathfrak{r}$-Untergruppe von G/K, so gibt es zu $o(G)_{\mathfrak{r}}$ teilerfremde ganze Zahlen e und f mit $X^e = Y$ und $Y^f = X$. [Um die Existenz von e und f einzusehen, braucht man sich nur zu überlegen, daß irgendzwei Elemente gleicher Ordnung in der zyklischen Gruppe der Ordnung $o(G)_{\mathfrak{r}}$ durch einen Automorphismus dieser zyklischen Gruppe ineinander übergeführt werden können, und daß Automorphismen dieser zyklischen Gruppe genau im Potenzieren mit einer zu $o(G)_{\mathfrak{r}}$ teilerfremden Zahl bestehen.] Da aber e und f zur Ordnung eines jeden $\mathfrak{r}$-Elements aus G teilerfremd sind, so bewirkt das Potenzieren mit e bzw. f eine Permutation der $\mathfrak{r}$-Elemente aus G. Es folgt sofort, daß die Restklassen X und Y von G modulo K gleichviel $\mathfrak{r}$-Elemente enthalten.

Wir wollen als nächstes zeigen:

(*) Ist X ein $\mathfrak{r}$-Element aus G/K, so ist die Anzahl der $\mathfrak{r}$-Elemente in der Restklasse X von G modulo K genau $o(K)_\mathfrak{r}$.

Diese Aussage (*) ist sicher richtig, wenn $o(X)=1$ ist, da dann ja $X=K$ ist. Wir können also die Induktionsannahme machen, daß die Aussage (*) für alle die $\mathfrak{r}$-Elemente aus G/K gelte, deren Ordnung kleiner als die Ordnung des $\mathfrak{r}$-Elements X ist. Wir bezeichnen mit a die Anzahl der $\mathfrak{r}$-Elemente in der Restklasse X. Die Zahl a ist sicher positiv. Dann bilden wir die Untergruppe $U=\{K, X\}$ von G und bemerken, daß U/K eine zyklische $\mathfrak{r}$-Gruppe der Ordnung $o(X)$ ist. Da $o(X)$ nur durch Primzahlen aus $\mathfrak{r}$ teilbar ist, so wird $o(U)_\mathfrak{r}=[o(X)o(K)]_\mathfrak{r}=o(X)o(K)_\mathfrak{r}$; und dies ist gerade die Anzahl der $\mathfrak{r}$-Elemente aus U, da ja G eine schwach $\mathfrak{r}$-abgeschlossene Gruppe ist. Ist nun Y ein Element aus U/K, so ist entweder Y ein erzeugendes Element von U/K; und dann enthält Y genauso viele $\mathfrak{r}$-Elemente wie X, nämlich a. Ist aber Y kein erzeugendes Element von U/K, so ist $o(Y)<o(X)$; und es folgt aus der Induktionsannahme, daß die Anzahl der $\mathfrak{r}$-Elemente in Y genau $o(K)_\mathfrak{r}$ ist. Ist die Anzahl der Elemente der Ordnung $o(X)$ in U/K etwa j, so folgt, daß die Zahl der $\mathfrak{r}$-Elemente in U genau

$$o(X)\,o(K)_\mathfrak{r}=o(U)_\mathfrak{r}=[o(X)-j]\,o(K)_\mathfrak{r}+j\,a$$

ist. Hieraus folgt aber $a=o(K)_\mathfrak{r}$, womit der induktive Beweis von (*) vollständig geführt ist.

Sei nun V/K irgendeine Untergruppe von G/K und v die Anzahl der $\mathfrak{r}$-Elemente aus V/K. Aus (*) folgt dann, daß die Zahl der $\mathfrak{r}$-Elemente in V genau $vo(K)_\mathfrak{r}$ ist. Da aber G eine schwach $\mathfrak{r}$-abgeschlossene Gruppe ist, so ist die Zahl der $\mathfrak{r}$-Elemente in V genau $o(V)_\mathfrak{r}$. Zusammenfassend ergibt sich

$$v=o(V)_\mathfrak{r}\,o(K)_\mathfrak{r}^{-1}=[o(V)\,o(K)^{-1}]_\mathfrak{r}=o(V/K)_\mathfrak{r}$$

womit die schwache $\mathfrak{r}$-Abgeschlossenheit von G/K bewiesen ist.

Für unsere Betrachtungen wird sich der von uns an anderer Stelle eingeführte Begriff der $\mathfrak{r}'$-*Homogenität* als nützlich erweisen. Unter einer $\mathfrak{r}'$-homogenen Gruppe verstehen wir eine Gruppe G derart, *daß* $\mathfrak{N}U/\mathfrak{C}U$ *eine* $\mathfrak{r}'$-*Gruppe für jede* $\mathfrak{r}'$-*Untergruppe* U *von* G *ist.* Man kann diese Eigenschaft etwas prägnanter, wenn auch weniger präzis, so ausdrücken: G induziert $\mathfrak{r}'$-Gruppen von Automorphismen in allen $\mathfrak{r}'$-Untergruppen von G. Untergruppen und Faktorgruppen $\mathfrak{r}'$-homogener Gruppen sind wieder $\mathfrak{r}'$-homogen [Baer, Lemma 2.3]. Daß jede $\mathfrak{r}$-abgeschlossene Gruppe auch $\mathfrak{r}'$-homogen ist, verifiziert man leicht; etwas tiefer liegt die folgende für uns wichtige Tatsache.

Lemma 2. *Jede schwach* $\mathfrak{r}$-*abgeschlossene Gruppe ist* $\mathfrak{r}'$-*homogen.*

Beweis. Wäre die schwach $\mathfrak{r}$-abgeschlossene Gruppe G nicht $\mathfrak{r}'$-homogen, so gäbe es eine $\mathfrak{r}'$-Untergruppe U und ein U normalisierendes Element, das in U einen Automorphismus induziert, der kein $\mathfrak{r}'$-Automorphismus ist. Dann gibt es auch ein U normalisierendes Element w, das in U einen Automorphismus von Primzahlordnung p aus $\mathfrak{r}$ induziert. Nun ist aber $w=ab$ das Produkt

zweier Potenzen a und b von w, von denen a ein p-Element ist, während $o(b)$ zu p teilerfremd ist. Dann induzieren a und w denselben Automorphismus in U und $V = \{U, a\} = U\{a\}$ ist eine Untergruppe der Ordnung $o(U)o(a)$ von G, da ja die p-Potenz $o(a)$ zu $o(U)$ teilerfremd ist. Da weiter p in $\mathfrak{r}$ liegt und U eine $\mathfrak{r}'$-Gruppe ist, so ist $o(V)_{\mathfrak{r}} = o(a)$. Aus der schwachen $\mathfrak{r}$-Abgeschlossenheit von G folgt dann, daß $o(a)$ die genaue Anzahl der $\mathfrak{r}$-Elemente aus V ist. Also müssen alle $\mathfrak{r}$-Elemente aus V in $\{a\}$ liegen, so daß $V_{\mathfrak{r}} = \{a\}$ eine charakteristische Untergruppe von V ist. Mithin ist $V = U \otimes \{a\}$ das direkte Produkt von U und $\{a\}$, so daß a mit allen Elementen aus U vertauschbar ist; und dies widerspricht der Wahl von a als Element, das in U einen Automorphismus der Ordnung p induziert. Aus diesem Widerspruch ergibt sich die $\mathfrak{r}'$-Homogenität der schwach $\mathfrak{r}$-abgeschlossenen Gruppen.

2. Aus Lemma 2 folgt, daß Satz A in folgendem Resultat als Spezialfall enthalten ist.

SATZ A*. *Die Gruppe G ist dann und nur dann $\mathfrak{r}$-abgeschlossen mit nilpotentem $G/G_{\mathfrak{r}}$, wenn G eine $\mathfrak{r}'$-homogene Gruppe ist, deren sämtliche $\mathfrak{r}'$-Untergruppen nilpotent sind.*

BEMERKUNG. Eine andere Charakterisierung dieser Gruppenklasse hat D. G. HIGMAN, p. 491, Theorem 4.3 angegeben. Es scheint nicht einfach zu sein, das Higmansche und das obige Kriterium direkt aus einander abzuleiten.

BEWEIS. Daß alle $\mathfrak{r}$-abgeschlossenen Gruppen auch $\mathfrak{r}'$-homogen sind, haben wir schon erwähnt. Ist G eine $\mathfrak{r}$-abgeschlossene Gruppe mit nilpotentem $G/G_{\mathfrak{r}}$, so ist jede $\mathfrak{r}'$-Untergruppe von G einer Untergruppe von $G/G_{\mathfrak{r}}$ isomorph und also nilpotent, womit die Notwendigkeit unserer Bedingungen dargetan ist.

Sei umgekehrt G eine $\mathfrak{r}'$-homogene Gruppe, deren $\mathfrak{r}'$-Untergruppen sämtlich nilpotent sind. Sei p ein nicht in $\mathfrak{r}$ liegender Primteiler von $o(G)$ und P irgendeine p-Untergruppe von G. Da P eine $\mathfrak{r}'$-Untergruppe von G ist, so folgt aus der $\mathfrak{r}'$-Homogenität von G, daß $\mathfrak{N}P/\mathfrak{C}P$ eine $\mathfrak{r}'$-Gruppe ist. Jedes Element X aus $\mathfrak{N}P/\mathfrak{C}P$ ist also ein $\mathfrak{r}'$-Element; und die Restklasse X enthält folglich auch ein $\mathfrak{r}'$-Element x. Da x im Normalisator der $\mathfrak{r}'$-Untergruppe P liegt, so ist $\{P, x\} = P\{x\}$ eine $\mathfrak{r}'$-Gruppe und als solche nilpotent. Nun ist $\{x\}$ das direkte Produkt einer p-Gruppe $\{a\}$ und einer Gruppe $\{b\}$ von zu p teilerfremder Ordnung mit $x = ab$. Da b in der nilpotenten Gruppe $\{P, x\}$ liegt, so ist b sogar in $\mathfrak{C}P$ enthalten; und es folgt, daß $X = (\mathfrak{C}P)x = (\mathfrak{C}P)a$ sogar ein p-Element ist. Also ist $\mathfrak{N}P/\mathfrak{C}P$ eine p-Gruppe; und damit haben wir sogar die p-Homogenität von G dargetan. Aus einem Satz von FROBENIUS folgt aber die Äquivalenz von p-Homogenität und p'-Abgeschlossenheit; s. etwa BAER, Theorem 5.1. Damit haben wir gezeigt: Zu jedem nicht in $\mathfrak{r}$ enthaltenen Primteiler p von $o(G)$ existiert eine charakteristische Untergruppe $G_{p'}$ von zu p teilerfremder Ordnung mit p-Faktorgruppe $G/G_{p'}$; und $G_{p'}$ ist genau die Gesamtheit der Elemente mit zu p teilerfremder Ordnung. Da aber ein Element dann und nur dann ein $\mathfrak{r}$-Element ist, wenn seine Ordnung zu jeder nicht in $\mathfrak{r}$ gelegenen Primzahl teilerfremd ist, so erweist sich die Menge $G_{\mathfrak{r}}$ genau als der Durchschnitt aller $G_{p'}$ mit p in $\mathfrak{r}'$. Also ist $G_{\mathfrak{r}}$ eine Untergruppe

von G, woraus die $\mathfrak{r}$-Abgeschlossenheit von G folgt. Da Ordnung $o(G_\mathfrak{r})$ und Index $[G:G_\mathfrak{r}]$ teilerfremd sind, so folgt aus dem Schurschen Satz die Existenz eines Komplements W von $G_\mathfrak{r}$ in G; s. etwa ZASSENHAUS, p. 125, Satz 25. Dann ist W zu $G/G_\mathfrak{r}$ isomorph. Also ist W eine $\mathfrak{r}'$-Untergruppe von G; und als solche ist W nilpotent. Die zu W isomorphe Gruppe $G/G_\mathfrak{r}$ ist mithin auch nilpotent, q.e.d.

3. BEWEIS *von Satz* B. Ist G eine $\mathfrak{r}$-abgeschlossene Gruppe mit nilpotentem $G_\mathfrak{r}$, so ist G, wie schon mehrfach bemerkt, schwach $\mathfrak{r}$-abgeschlossen und jede $\mathfrak{r}$-Untergruppe von G ist in $G_\mathfrak{r}$ enthalten und also nilpotent. Damit ist die Notwendigkeit unserer Bedingungen dargetan.

Wären unsere Bedingungen nicht hinreichend, so gäbe es eine Gruppe E kleinster Ordnung mit den folgenden drei Eigenschaften:

(1) E ist schwach $\mathfrak{r}$-abgeschlossen.

(2) Jede $\mathfrak{r}$-Untergruppe von E ist nilpotent.

(3) E ist nicht $\mathfrak{r}$-abgeschlossen.

[Wäre nämlich E eine $\mathfrak{r}$-abgeschlossene Gruppe, so würde ja aus (2) sofort auch die Nilpotenz der $\mathfrak{r}$-Untergruppe $E_\mathfrak{r}$ folgen.]

Ist V eine echte Untergruppe von E, so ist V mit E eine schwach $\mathfrak{r}$-abgeschlossene Gruppe, deren $\mathfrak{r}$-Untergruppen wegen (2) nilpotent sind. Aus der Minimalität von E folgt also:

(4) Ist V eine echte Untergruppe von F, so ist V eine $\mathfrak{r}$-abgeschlossene Gruppe mit nilpotentem $V_\mathfrak{r}$.

Aus Lemma 2 und der schwachen $\mathfrak{r}$-Abgeschlossenheit von E folgern wir:

(5) E ist $\mathfrak{r}'$-homogen.

Wäre E nicht einfach, so gäbe es einen $K<E$ erfüllenden minimalen Normalteiler K von E. Aus (4) folgt die $\mathfrak{r}$-Abgeschlossenheit von K, so daß die Gesamtheit der $\mathfrak{r}$-Elemente aus K eine charakteristische $\mathfrak{r}$-Untergruppe $K_\mathfrak{r}$ ist. Charakteristische Untergruppen von Normalteilern sind Normalteiler. Also ist $K_\mathfrak{r}$ ein Normalteiler von E; und aus der Minimalität von K folgt, daß $K_\mathfrak{r}=1$ oder $=K$ ist. Im ersten Falle ist K eine $\mathfrak{r}'$-Gruppe und im zweiten eine $\mathfrak{r}$-Gruppe. — Aus Lemma 1 und der schwachen $\mathfrak{r}$-Abgeschlossenheit von E folgt die schwache $\mathfrak{r}$-Abgeschlossenheit von E/K. Ist weiter U/K eine $\mathfrak{r}$-Untergruppe von E/K, so ist U eine $\mathfrak{r}$-Untergruppe von E, wenn K eine $\mathfrak{r}$-Gruppe ist. In diesem Falle ist U als $\mathfrak{r}$-Untergruppe von E wegen (2) nilpotent; und hieraus folgt die Nilpotenz von U/K. Ist aber K keine $\mathfrak{r}$-Gruppe, so ist K eine $\mathfrak{r}'$-Gruppe, wie gerade gezeigt. Dann sind $o(K)$ und $[U:K]$ teilerfremd; und wir erschließen aus dem Schurschen Satz die Existenz eines Komplements V von K in U; s. ZASSENHAUS, p. 125, Satz 25. Aus der Isomorphie von V und U/K folgt zunächst, daß V eine $\mathfrak{r}$-Gruppe ist, daß also die $\mathfrak{r}$-Untergruppe V von E wegen (2) nilpotent ist, und daß schließlich die isomorphe Gruppe U/K ebenfalls nilpotent ist. Damit haben wir ganz allgemein die Nilpotenz aller $\mathfrak{r}$-Untergruppen von E/K dargetan. Da E/K kleinere

Ordnung als E hat, so folgt jetzt aus der Minimalität von E die $\mathfrak{r}$-Abgeschlossenheit von E/K. Die nach (5) $\mathfrak{r}'$-homogene Erweiterung E der $\mathfrak{r}$-abgeschlossenen Gruppe K durch die $\mathfrak{r}$-abgeschlossene Gruppe E/K ist dann aber auch $\mathfrak{r}$-abgeschlossen, wie wir an anderer Stelle gezeigt haben; BAER, Lemma 2.4. Dies widerspricht (3); und damit haben wir gezeigt:

(6) E ist einfach.

Wir betrachten eine maximale Untergruppe A von E, deren Ordnung durch Primzahlen aus $\mathfrak{r}$ teilbar ist. Dann ist $o(A)_\mathfrak{r} \neq 1$; und wir erschließen aus (4) die Existenz der nilpotenten charakteristischen $\mathfrak{r}$-Untergruppe $A_\mathfrak{r}$ von A, deren Ordnung $o(A_\mathfrak{r}) = o(A)_\mathfrak{r}$ ist. Also ist auch $1 < A_\mathfrak{r} \leqq A < E$; und es folgt aus (6), daß $A_\mathfrak{r}$ kein Normalteiler von E ist. Mithin wird $A \leqq \mathfrak{N}(A_\mathfrak{r}) < E$, woraus $A = \mathfrak{N}(A_\mathfrak{r})$ wegen der Maximalität von A folgt. — Sei weiter R eine $A_\mathfrak{r}$ enthaltende $\mathfrak{r}$-Untergruppe von E. Wegen (2) ist R nilpotent. Wäre also $A_\mathfrak{r} < R$, so folgte aus den bekannten Eigenschaften der nilpotenten Gruppen, daß der Normalisator von $A_\mathfrak{r}$ in R von $A_\mathfrak{r}$ verschieden ist. Es wäre also $A_\mathfrak{r} < R \cap \mathfrak{N}(A_\mathfrak{r}) = R \cap A \leqq A_\mathfrak{r}$, da ja R eine $\mathfrak{r}$-Gruppe ist und $A_\mathfrak{r}$ alle $\mathfrak{r}$-Elemente aus A enthält. Dies ist unmöglich; und folglich ist $A_\mathfrak{r}$ eine maximale $\mathfrak{r}$-Untergruppe von E. Damit haben wir folgendes bewiesen:

(7) Ist die maximale Untergruppe A von E keine $\mathfrak{r}'$-Gruppe, so ist $A_\mathfrak{r}$ eine maximale $\mathfrak{r}$-Untergruppe von E und $A = \mathfrak{N}(A_\mathfrak{r})$.

Sei jetzt B eine maximale $\mathfrak{r}$-Untergruppe von E. Da E wegen (3) weder eine $\mathfrak{r}$-Gruppe noch eine $\mathfrak{r}'$-Gruppe ist, so ist $1 < B < E$ und B ist also in einer maximalen Untergruppe A von E enthalten, die keine $\mathfrak{r}'$-Gruppe ist. Aus (7) folgt $A = \mathfrak{N}(A_\mathfrak{r})$, wo die charakteristische $\mathfrak{r}$-Untergruppe $A_\mathfrak{r}$ von A genau die Gesamtheit der $\mathfrak{r}$-Elemente aus A ist. Also ist $B \leqq A_\mathfrak{r}$ und aus der Maximalität von B folgt sogar $B = A_\mathfrak{r}$, so daß $A = \mathfrak{N}B$ wird. Damit haben wir folgendes bewiesen:

(8) Ist B eine maximale $\mathfrak{r}$-Untergruppe von E, so ist $\mathfrak{N}B$ die einzige B enthaltende maximale Untergruppe von E; und es wird $B = (\mathfrak{N}B)_\mathfrak{r}$.

Gäbe es verschiedene maximale $\mathfrak{r}$-Untergruppen von E, deren Durchschnitt von 1 verschieden ist, so gäbe es auch ein solches Paar U, V verschiedener maximaler $\mathfrak{r}$-Untergruppen von E, deren Durchschnitt $D = U \cap V \neq 1$ maximale Ordnung [unter allen diesen Durchschnitten] hätte. Aus der Maximalität von U und V und aus ihrer Verschiedenheit folgt $D < U$ und $D < V$. Aus (2) ergibt sich die Nilpotenz von U und V. Aus den wohlbekannten Eigenschaften nilpotenter Gruppen folgt weiter $D < U \cap \mathfrak{N}D$ und $D < V \cap \mathfrak{N}D$. Aus $1 < D < E$ und (6) ergibt sich, daß D kein Normalteiler von E und also $\mathfrak{N}D < E$ ist. Also ist $\mathfrak{N}D$ in einer maximalen Untergruppe A von E enthalten, die keine $\mathfrak{r}'$-Gruppe ist, da sie die $\mathfrak{r}$-Gruppe $D \neq 1$ enthält. Aus (7) folgt dann, daß die Gesamtheit $A_\mathfrak{r}$ der $\mathfrak{r}$-Elemente aus A eine maximale $\mathfrak{r}$-Untergruppe von E ist, und daß $A = \mathfrak{N}(A_\mathfrak{r})$ gilt. Folglich wird

$$D < U \cap \mathfrak{N}D \leqq U \cap A = U \cap A_\mathfrak{r},$$

da U eine $\mathfrak{r}$-Gruppe ist; und aus der Maximalität des Durchschnitts D der maximalen $\mathfrak{r}$-Untergruppen U und V erschließen wir $U = A_\mathfrak{r}$. Ebenso sieht

man $A_\mathfrak{r} = V$ ein, woraus der Widerspruch $U = V$ folgt. Damit haben wir gezeigt:

(9) Sind U und V verschiedene maximale $\mathfrak{r}$-Untergruppen von E, so ist $U \cap V = 1$.

Ist B eine maximale $\mathfrak{r}$-Untergruppe von E, so ist $A = \mathfrak{N}B$ wegen (8) die einzige B enthaltende maximale Untergruppe von E; und es ist $B = A_\mathfrak{r}$ die Gesamtheit der $\mathfrak{r}$-Elemente aus A. Also ist A/B eine $\mathfrak{r}'$-Gruppe; und folglich wird $[A:B]_\mathfrak{r} = 1$ und mithin

$$o(E)_\mathfrak{r} = [E:A]_\mathfrak{r} o(B).$$

Wegen (1) ist $o(E)_\mathfrak{r}$ die genaue Anzahl der $\mathfrak{r}$-Elemente in E. Da weiter $[E:\mathfrak{N}B] = [E:A]$ die Anzahl der zu B konjugierten Untergruppen ist, und da zu B konjugierte Untergruppen als maximale $\mathfrak{r}$-Untergruppen wegen (9) paarweise nur die 1 gemein haben, so ist $[o(B)-1][E:A]$ die Anzahl der von 1 verschiedenen Elemente, die in zu B konjugierten Untergruppen liegen. Da diese sämtlich $\mathfrak{r}$-Elemente sind, so folgt

$$1 + [o(B) - 1][E:A] \leqq o(E)_\mathfrak{r} = [E:A]_\mathfrak{r} o(B)$$

oder

$$o(B)[E:A] < o(B)[E:A]_\mathfrak{r} + [E:A].$$

Setzen wir $[E:A] = h[E:A]_\mathfrak{r}$, so ist h eine positive ganze Zahl, die der Ungleichung

$$o(B)h < o(B) + h$$

genügt. Diese Ungleichung ist mit

$$[o(B) - 1][h - 1] < 1$$

gleichwertig. Aus $1 < B$ folgt aber $0 < o(B) - 1$; und hieraus ergibt sich schließlich $h = 1$, so daß $[E:A] = [E:A]_\mathfrak{r}$ wird. Damit haben wir folgendes gezeigt:

(10) Ist B eine maximale $\mathfrak{r}$-Untergruppe von E, so ist $[E:\mathfrak{N}B]$ nur durch Primzahlen aus $\mathfrak{r}$ teilbar und es ist $o(E)_\mathfrak{r} = [E:\mathfrak{N}B] o(B)_\mathfrak{r}$.

Die maximalen $\mathfrak{r}$-Untergruppen von E sind sämtlich von 1 und E verschieden. Sie sind in j Klassen $\mathfrak{K}_1, \ldots, \mathfrak{K}_j$ zueinander konjugierter maximaler $\mathfrak{r}$-Untergruppen von E aufgeteilt. Alle Untergruppen in der Klasse $\mathfrak{K}_i$ haben die gleiche Ordnung k_i; und die Anzahl der Untergruppen in der Klasse $\mathfrak{K}_i$ werde mit a_i bezeichnet. Aus (10) folgt:

$$o(E)_\mathfrak{r} = k_i a_i \quad \text{für } i = 1, \ldots, j.$$

Da jedes von 1 verschiedene $\mathfrak{r}$-Element wegen (9) in einer und nur einer maximalen $\mathfrak{r}$-Untergruppe von E enthalten ist, so ist die Anzahl der $\mathfrak{r}$-Elemente in E wegen (1) genau

$$(*) \qquad o(E)_\mathfrak{r} = 1 + \sum_{i=1}^{j} a_i(k_i - 1) = 1 + j\, o(E)_\mathfrak{r} - \sum_{i=1}^{j} a_i.$$

Da alle maximalen $\mathfrak{r}$-Untergruppen von E mindestens die Ordnung 2 haben, so wird $2 \leqq k_i$ und also $2a_i \leqq k_i a_i = o(E)_\mathfrak{r}$, woraus wir $a_i \leqq \frac{1}{2} o(E)_\mathfrak{r}$ erschließen. Also wird:

$$1 + (j-1)\, o(E)_\mathfrak{r} = \sum_{i=1}^{j} a_i \leqq \tfrac{1}{2} o(E)_\mathfrak{r} j;$$

und dies ist mit

$$1 \leqq o(E)_\mathfrak{r}(1 - \tfrac{1}{2} j)$$

gleichwertig. Dann muß aber $1 - \frac{1}{2} j$ positiv sein, woraus $0 < j < 2$ folgt. Also ist $j = 1$. Aus (*) folgt weiter

$$1 = a_1,$$

so daß die einzige Klasse $\mathfrak{K}_1$ nur aus einer einzigen Untergruppe besteht. Diese ist eine maximale $\mathfrak{r}$-Untergruppe und wäre also ein echter Normalteiler von E im Widerspruch zu (6). Aus diesem Widerspruch folgt die Gültigkeit von Satz B.

4. BEWEIS *von Satz* C. Ist G eine $\mathfrak{r}$-abgeschlossene Gruppe mit auflösbarem $G/G_\mathfrak{r}$, so ist G natürlich auch schwach $\mathfrak{r}$-abgeschlossen. Ist E eine einfache Faktorgruppe einer Untergruppe von G, so ist E ebenfalls eine $\mathfrak{r}$-abgeschlossene Gruppe mit auflösbarem $E/E_\mathfrak{r}$. Aus der Einfachheit von E folgt $E_\mathfrak{r} = 1$ oder $E_\mathfrak{r} = E$. Im ersten Falle ist E eine einfache auflösbare $\mathfrak{r}'$-Gruppe, deren Ordnung natürlich eine in $\mathfrak{r}'$ enthaltene Primzahl ist; und im zweiten Falle ist E eine $\mathfrak{r}$-Gruppe. In beiden Fällen ist $o(E)$ durch höchstens eine Primzahl aus $\mathfrak{r}'$ teilbar. Damit ist die Notwendigkeit der Bedingungen des Satzes C dargetan.

Sei H irgendeine $\mathfrak{r}$-abgeschlossene Gruppe derart, daß $o(E)$ für jede einfache Faktorgruppe einer Untergruppe von H durch höchstens eine Primzahl aus $\mathfrak{r}'$ teilbar ist. Dann ist jeder Kompositionsfaktor der $\mathfrak{r}'$-Gruppe $H/H_\mathfrak{r}$ eine p-Gruppe, so daß $H/H_\mathfrak{r}$ auflösbar ist. Zum Nachweis des Hinreichens der Bedingungen des Satzes C genügt also der Beweis der $\mathfrak{r}$-Abgeschlossenheit aller diese Bedingungen erfüllenden Gruppen.

Wären die Bedingungen des Satzes C nicht hinreichend, so gäbe es nach dem soeben Bemerkten eine Gruppe E kleinster Ordnung mit den folgenden drei Eigenschaften:

(1) E ist schwach $\mathfrak{r}$-abgeschlossen.

(2) Ist F eine einfache Faktorgruppe einer Untergruppe von E, so ist $o(F)$ durch höchstens eine Primzahl aus $\mathfrak{r}'$ teilbar.

(3) E ist nicht $\mathfrak{r}$-abgeschlossen.

Aus Lemma 1 folgt, daß sich Bedingung (1) auf Untergruppen und Faktorgruppen vererbt; und die Erblichkeit der Bedingung (2) ist bereits in ihrer Formulierung enthalten. Aus der Minimalität von E folgt also:

(4) Jede echte Untergruppe und jede echte Faktorgruppe von E ist $\mathfrak{r}$-abgeschlossen.

Wäre E nicht einfach, so gäbe es einen $1 < K < E$ erfüllenden Normalteiler K von E. Aus (4) folgt die $\mathfrak{r}$-Abgeschlossenheit von K und E/K. Aus Lemma 2

folgt die $\mathfrak{r}'$-Homogeneität der schwach $\mathfrak{r}$-abgeschlossenen Gruppe E. Nun sind $\mathfrak{r}'$-homogene Erweiterungen von $\mathfrak{r}$-abgeschlossenen Gruppen durch $\mathfrak{r}$-abgeschlossene Gruppen selbst $\mathfrak{r}$-abgeschlossen; s. BAER, Lemma 2.4. Damit wäre aber die (3) widersprechende $\mathfrak{r}$-Abgeschlossenheit von E bewiesen. Es folgt:

(5) E ist einfach.

Aus (3) folgt insbesondere, daß E weder eine $\mathfrak{r}$-Gruppe noch eine $\mathfrak{r}'$-Gruppe ist. Aus (2) folgt aber wegen der Einfachheit von E, daß $o(E)$ durch höchstens eine Primzahl aus $\mathfrak{r}'$ teilbar ist. Damit haben wir bewiesen:

(6) Es gibt einen und nur einen in $\mathfrak{r}'$ enthaltenen Primteiler $\mathfrak{p}$ von $o(E)$.

Ist $P \neq 1$ eine $\mathfrak{p}$-Untergruppe von E, so ist P wegen (5) kein Normalteiler von E. Also ist $\mathfrak{N}P \neq E$; und es folgt aus (4), daß $\mathfrak{N}P$ eine $\mathfrak{r}$-abgeschlossene Gruppe ist. Da aber $\mathfrak{p}$ der einzige nicht in $\mathfrak{r}$ enthaltene Primteiler von $o(E)$ ist, so ist $\mathfrak{N}P/(\mathfrak{N}P)_\mathfrak{r}$ eine $\mathfrak{p}$-Gruppe. Da P ein Normalteiler von $\mathfrak{N}P$ und $P \cap (\mathfrak{N}P)_\mathfrak{r} = 1$ ist, so ist $(\mathfrak{N}P)_\mathfrak{r}$ sogar im Zentralisator von P enthalten. Es folgt, daß $\mathfrak{N}P/\mathfrak{C}P$ eine $\mathfrak{p}$-Gruppe ist. Damit haben wir bewiesen, daß E eine $\mathfrak{p}$-homogene Gruppe ist. Eine solche besitzt aber nach einem Satz von FROBENIUS einen Normalteiler von zu $\mathfrak{p}$ teilerfremder Ordnung mit $\mathfrak{p}$-Faktorgruppe; s. BAER, Theorem 5.1. Aus (5) und (6) folgt dann aber, daß E eine $\mathfrak{p}$-Gruppe ist; und dies widerspricht (3). Aus diesem Widerspruch folgt die Gültigkeit von Satz C.

5. BEWEIS *von Satz* D. Ist G eine $\mathfrak{r}$-abgeschlossene Gruppe mit auflösbarem $G_\mathfrak{r}$, so ist G natürlich auch schwach $\mathfrak{r}$-abgeschlossen. Ist weiter E eine einfache Faktorgruppe einer Untergruppe von G, so ist auch E eine $\mathfrak{r}$-abgeschlossene Gruppe mit auflösbarem $E_\mathfrak{r}$. Aus der Einfachheit von E folgt $E_\mathfrak{r} = 1$ oder $E_\mathfrak{r} = E$. Im ersten Falle ist E eine $\mathfrak{r}'$-Gruppe und im zweiten Falle ist E eine einfache auflösbare Gruppe und also eine $\mathfrak{p}$-Gruppe. In beiden Fällen ist $o(E)$ durch höchstens eine Primzahl aus $\mathfrak{r}$ teilbar.

Sei H irgendeine $\mathfrak{r}$-abgeschlossene Gruppe derart, daß $o(E)$ für jede einfache Faktorgruppe E einer Untergruppe von H durch höchstens eine Primzahl aus $\mathfrak{r}$ teilbar ist. Dann ist jeder Kompositionsfaktor der $\mathfrak{r}$-Gruppe $H_\mathfrak{r}$ eine $\mathfrak{p}$-Gruppe, so daß $H_\mathfrak{r}$ auflösbar ist. Zum Nachweis des Hinreichens der Bedingungen von Satz D genügt also der Beweis der $\mathfrak{r}$-Abgeschlossenheit aller diese Bedingungen erfüllenden Gruppen.

Wären die Bedingungen von Satz D nicht hinreichend, so gäbe es nach dem soeben Bemerkten eine Gruppe E kleinster Ordnung mit folgenden Eigenschaften:

(1) E ist schwach $\mathfrak{r}$-abgeschlossen.

(2) Ist F eine einfache Faktorgruppe einer Untergruppe von E, so ist $o(F)$ durch höchstens eine Primzahl aus $\mathfrak{r}$ teilbar.

(3) E ist nicht $\mathfrak{r}$-abgeschlossen.

Aus Lemma 1 und der Minimalität von E ergibt sich zunächst:

(4) Jede echte Untergruppe und jede echte Faktorgruppe von E ist $\mathfrak{r}$-abgeschlossen.

Wäre E nicht einfach, so gäbe es einen $1<K<E$ erfüllenden Normalteiler K von E. Aus (4) folgt die $\mathfrak{r}$-Abgeschlossenheit von K und E/K; und aus Lemma 2 folgern wir die $\mathfrak{r}'$-Homogenität der schwach $\mathfrak{r}$-abgeschlossenen Gruppe E. Hieraus ergibt sich die $\mathfrak{r}$-Abgeschlossenheit von E; s. BAER, Lemma 2.4. Dies widerspricht (3). Also gilt:

(5) E ist einfach.

Aus (3) folgt insbesondere, daß E weder eine $\mathfrak{r}$-Gruppe noch eine $\mathfrak{r}'$-Gruppe ist. Aus (2) folgt aber wegen der Einfachheit von E, daß $o(E)$ durch höchstens eine Primzahl aus $\mathfrak{r}$ teilbar ist. Damit haben wir gezeigt:

(6) Ein und nur ein Primteiler $\mathfrak{p}$ von $o(E)$ gehört zu $\mathfrak{r}$.

Folglich ist $o(E)_{\mathfrak{r}}=o(E)_{\mathfrak{p}}$; und aus der schwachen $\mathfrak{r}$-Abgeschlossenheit von E ergibt sich, daß $o(E)_{\mathfrak{r}}=o(E)_{\mathfrak{p}}$ die Anzahl der $\mathfrak{r}$-Elemente in E ist. Nun enthält aber eine $\mathfrak{p}$-Sylowuntergruppe P von E genau $o(E)_{\mathfrak{p}}$ Elemente. Es folgt, daß P die Gesamtheit $E_{\mathfrak{r}}$ der $\mathfrak{r}$-Elemente aus E ist. Folglich ist $P\neq 1$ eine charakteristische Untergruppe von E. Aus der Einfachheit von E — s. (5) — folgt $E=P$, so daß E eine $\mathfrak{p}$-Gruppe und also $\mathfrak{r}$-abgeschlossen ist. Dies widerspricht (3); und aus diesem Widerspruch folgt Satz D.

Literatur

BAER, R.: Closure and Dispersion. Illinois J. Math. 2, 619—640 (1958). — FROBENIUS, G.: Über auflösbare Gruppen IV, V. Sitzgsber. preuß. Akad. Wiss. 1216 bis 1230, 1324—1329 (1901). — HIGMAN, D. G.: Focal Series in Finite Groups. Canad. J. Math. 5, 477—497 (1953). — ZASSENHAUS, H.: Lehrbuch der Gruppentheorie. Leipzig und Berlin 1937.

Frankfurt a. M., Mathematisches Seminar der Universität

(Eingegangen am 19. Februar 1959)

LEICHTWEISS, K.
Math. Zeitschr. 71, 335–360 (1959).

Selbstadjungierte Banach-Räume

Von
KURT LEICHTWEISS

§ 1. Einleitung

Das Problem der Kennzeichnung des euklidischen Raums bzw. des (reellen) Hilbert-Raums unter den Banach-Räumen ist in der Literatur schon wiederholt untersucht worden. So zeigten z. B. P. JORDAN und J. v. NEUMANN [1], daß ein reeller, separabler Banach-Raum, in welchem für je zwei Punkte x und y die Beziehung $\|x+y\|^2+\|x-y\|^2=2(\|x\|^2+\|y\|^2)$ besteht, notwendig ein euklidischer Raum oder ein Hilbert-Raum sein muß. Diese Gleichung ist später stark abgeschwächt worden, und man hat ähnliche Bedingungen gefunden[1]). R. C. JAMES [3] gab, auf ein Resultat G. BIRKHOFFs zurückgreifend, als eine andere, für den Hilbert-Raum sowie für den euklidischen Raum mit einer Dimension größer als zwei charakteristische Eigenschaft die Tatsache an, daß in diesen Räumen die Transversalität zweier Vektoren eine symmetrische Relation darstellt[2]). JAMES nennt dabei den zum Punkte x gehörenden (Orts)vektor transversal zum entsprechenden Vektor von y, wenn für alle reellen Skalare α $\|x+\alpha y\| \geqq \|x\|$ gilt. (J. RADON [5] hatte schon lange vorher gezeigt, daß ein linearer normierter Raum der Dimension zwei mit symmetrischer Transversalität nicht unbedingt eine euklidische Ebene ist, so daß die Dimension zwei hierbei wirklich einen Ausnahmefall darstellt.)

Von E. R. LORCH [6] stammt eine weitere bemerkenswerte Arbeit in der angegebenen Richtung, woran wir im folgenden anknüpfen werden. LORCH geht von der Tatsache aus, daß beim euklidischen Raum und beim Hilbert-Raum der Dualraum $\mathfrak{L}^*$ mit dem Raum $\mathfrak{L}$ in genau derselben Weise identifiziert werden kann, wie dies bei einem reflexiven linearen normierten Raum $\mathfrak{L}$ für den Bidualraum $\mathfrak{L}^{**}$ und $\mathfrak{L}$ möglich ist. Er definiert nun für einen allgemeinen linearen normierten Raum $\mathfrak{L}$ die Gleichheit von $\mathfrak{L}^*$ und $\mathfrak{L}$ durch die Forderung der Existenz eines linearen Homöomorphismus A der Norm 1 von $\mathfrak{L}^*$ auf $\mathfrak{L}$, für den die Gleichung $f(Af)=\|f\|\,\|Af\|$ für alle $f\in\mathfrak{L}^*$ erfüllt ist ($f(Af)$ ist der Wert des Funktionals f an der Stelle $Af\in\mathfrak{L}$), und beweist: Für jeden separablen reellen Banach-Raum $\mathfrak{L}$ mit $\mathfrak{L}^*=\mathfrak{L}$ gilt $\|Af\|=\|f\|$, und dieser ist entweder ein euklidischer Raum oder ein Hilbert-Raum[3]). A bewirkt also insbesondere einen Isomorphismus, d. h. eine lineare und die Norm erhaltende Abbildung von $\mathfrak{L}^*$ auf $\mathfrak{L}$. Die naheliegende Frage, ob schon die

[1]) Viele derartige Bedingungen nebst der zugehörigen Literatur findet man bei LORCH [2].

[2]) Dies wurde wie viele andere diesbezügliche Sätze auch von M. M. DAY [4] bewiesen.

[3]) Die hier angegebenen Charakterisierungen des euklidischen Raums und des Hilbert-Raums findet man auch als Übungen bei N. BOURBAKI [7], V, § 1, exerc.

Isomorphie von $\mathfrak{L}^*$ und $\mathfrak{L}$ allein für den euklidischen Raum bzw. den Hilbert-Raum kennzeichnend ist, läßt sich allerdings an Hand eines Gegenbeispiels verneinen[4]), so daß aus diesem Grunde die linearen normierten Räume $\mathfrak{L}$ mit (mindestens) einem Isomorphismus A von $\mathfrak{L}^*$ auf $\mathfrak{L}$ oder — wie wir dieselben kurz nennen wollen — die „selbstadjungierten" linearen normierten Räume eine Mittelstellung zwischen den allgemeinen linearen normierten Räumen und dem euklidischen Raum bzw. dem Hilbert-Raum einnehmen.

Ziel der vorliegenden Arbeit ist nun eine genauere Untersuchung dieser selbstadjungierten Räume. Dabei wird ein etwas verallgemeinerter Begriff der Norm zugrunde gelegt, welcher sich von dem üblichen dadurch unterscheidet, daß man an Stelle von $\|\alpha x\| = |\alpha| \, \|x\|$ für alle (reellen) α nur $\|\alpha x\| = \alpha \|x\|$ für $\alpha \geqq 0$ postuliert. Dies macht zunächst die Neufassung wichtiger Definitionen notwendig (§ 2), läßt sich aber ohne weiteres durchführen, da die Topologie des betrachteten Raums im Zusammenhang mit unseren Überlegungen gar keine Rolle spielt. Daraufhin lassen sich einige allgemeine Sätze mehr formaler Art für die „schiefnormierten" selbstadjungierten Räume beliebiger Dimension ableiten (§ 3). Von besonderer Bedeutung erweist sich dabei die Tatsache, daß bei einem schiefnormierten selbstadjungierten und reflexiven Raum $\mathfrak{L}$ mit dem Isomorphismus A aus A die einen Automorphismus von $\mathfrak{L}$ darstellende Abbildung $A(A^*)^{-1}$ gebildet werden kann (Satz 6). Dadurch wird es nämlich möglich — und hier beginnen die wesentlich neuen Resultate der Arbeit —, unter allen Isomorphismen A eines endlichdimensionalen schiefnormierten selbstadjungierten Raumes immer einen Isomorphismus von einer besonders einfachen, leicht angebbaren Form zu finden (Satz 10), die durch geeignete Beispiele realisiert werden kann (§ 4). In § 5 werden weiter Kriterien dafür angegeben, wann ein (schiefnormierter) selbstadjungierter Minkowski-Raum ein euklidischer Raum ist; diese sind unter anderem von arithmetischer und von geometrischer Natur (Satz 13 und Satz 14). Die explizite Bestimmung der Norm aller selbstadjungierten Minkowski-Räume, welche durch die Ergebnisse des § 4 vorbereitet wird, bleibt selbst einer zukünftigen Arbeit vorbehalten.

Bezüglich der Bezeichnungsweise bemerken wir, daß Punkte eines linearen schiefnormierten Raums mit $x, y, \ldots$ und Skalare mit $\alpha, \beta, \ldots$ bezeichnet werden; Räume mit komplexen Skalaren bleiben von unseren Betrachtungen ausgeschlossen.

§ 2. Definitionen

Es ist für das Folgende zweckmäßig, den üblichen Begriff eines linearen normierten Raums in der folgenden Weise etwas zu verallgemeinern:

DEFINITION 1. *Ein Raum $\mathfrak{L}$ heiße (reeller) linearer schiefnormierter Raum, wenn er ein linearer Raum mit reellen Skalaren ist und wenn er schiefnormiert ist, d.h. wenn jedem Punkt $x \in \mathfrak{L}$ eine nichtnegative reelle Zahl $\|x\|$ mit den folgenden Eigenschaften zugeordnet ist:*

[4]) Vgl. dazu [6], S. 473.

a) Aus $||x|| = 0$ *folgt* $x = 0$,
b) $||\alpha x|| = \alpha\,||x||$ *falls* $\alpha \geqq 0$, *und*
c) $||x + y|| \leqq ||x|| + ||y||$ [5]).

Hierbei unterscheidet sich nur Axiom b) von den Axiomen eines (reellen) *linearen normierten Raums*, für den man bekanntlich $||\alpha x|| = |\alpha|\,||x||$ für alle α, also etwas mehr postuliert. Wir bemerken, daß Axiom c) gleichbedeutend mit der Konvexität der durch $||x|| \leqq \alpha$ definierten Punktmenge von $\mathfrak{L}$ ist.

In einem linearen schiefnormierten Raum $\mathfrak{L}$ läßt sich in einfacher Weise durch die Beziehung $|||x||| = \mathrm{Max}(||x||, ||-x||)$ eine neue Norm einführen. $\mathfrak{L}$ ist bezüglich dieser Norm ein linearer normierter Raum; und es ist wohlbekannt, daß die lineare Struktur und die durch die Norm $|||\ |||$ induzierte Topologie von $\mathfrak{L}$ kompatibel sind.

Definition 2. *Ein linearer schiefnormierter Raum heißt vollständig oder ein schiefer Banach-Raum, wenn er in bezug auf die Norm* $|||\ |||$ *vollständig ist.*

Definition 3. *Ein linearer schiefnormierter Raum heißt separabel, wenn er eine abzählbare und bezüglich der Normtopologie der Norm* $|||\ |||$ *überall dichte Punktmenge enthält.*

Definition 4. *Ein linearer schiefnormierter Raum von endlicher Dimensionszahl wird als Minkowski-Raum bezeichnet* (vgl. T. Bonnesen u. W. Fenchel: „Theorie der konvexen Körper", S. 23!).

Jeder Minkowski-Raum ist vollständig und separabel; für ihn stimmt die gewöhnliche Topologie des affinen Raums mit der Normtopologie von $|||\ |||$ (und mit der Normtopologie von $||\ ||$) überein.

In der Theorie der linearen normierten Räume betrachtet man unter anderem die linearen und beschränkten Funktionale eines solchen Raums und beweist, daß sie einen Raum mit denselben Eigenschaften, nämlich den sog. Dualraum bilden. Wir verallgemeinern dies für unsere linearen schiefnormierten Räume wie folgt:

Definition 5. *Ein (reelles) lineares Funktional* f *des linearen schiefnormierten Raumes* $\mathfrak{L}$ *werde beschränkt genannt, wenn eine solche positive Zahl* μ *existiert, daß* $f(x) \leqq \mu$ *für alle* $x \in \mathfrak{L}$ *mit* $||x|| = 1$ *ist.*

Jedes lineare Funktional eines Minkowski-Raums ist beschränkt, da in diesem Fall die Punktmenge $||x|| = 1$ kompakt ist. Jetzt gilt:

Satz 1: *Die Menge aller (reellen) beschränkten linearen Funktionale eines linearen schiefnormierten Raumes* $\mathfrak{L}$ *ist selbst ein linearer schiefnormierter Raum* $\mathfrak{L}^*$, *wenn man darin Addition, Skalarmultiplikation und Norm der Funktionale durch die Beziehungen:* $(f+g)(x) = f(x) + g(x)$; $(\alpha f)(x) = \alpha f(x)$ *und* $||f|| = \sup_{||x||=1} f(x)$ *erklärt.* $\mathfrak{L}^*$ *wird Dualraum von* $\mathfrak{L}$ *genannt.*

Beweis. Da $f \in \mathfrak{L}^*$ im Sinne von Definition 5 beschränkt ist, folgt die Endlichkeit von $||f||$. Diese Zahl ist sicher nicht negativ, weil sonst für ein

[5]) Lorch [8] betrachtete ebenfalls Räume dieses Typs und bewies die nachfolgenden Sätze 1, 2 und 3 (allerdings unter der dafür unwesentlichen, zusätzlichen Voraussetzung $||-x|| \leqq \gamma||x||$ für alle $x \in \mathfrak{L}$ und eine geeignete Zahl γ).

beliebiges $x \neq 0$ $f\left(\frac{x}{\|x\|}\right) = \frac{1}{\|x\|} f(x)$ und $f\left(\frac{-x}{\|-x\|}\right) = -\frac{1}{\|-x\|} f(x)$ gleichzeitig negativ wären, was offensichtlich unmöglich ist. $\|f\| = 0$ hat auf diese Weise $f(x) \equiv 0$, d.h. $f = 0$ zur Folge, und man verifiziert sofort: $\|\alpha f\| = \alpha \|f\|$ für $\alpha \geqq 0$ sowie $\|f+g\| \leqq \|f\| + \|g\|$. Wir bemerken noch, daß unsere Definition der Norm eines beschränkten Funktionals im Falle eines linearen normierten Raums $\mathfrak{L}$ mit der üblichen Definition übereinstimmt; w.z.b.w.

Das folgende Ergebnis stammt (in einer etwas anderen Formulierung) schon von BANACH ([9], chap. II, théorème 1):

Satz 2: *Ein auf einem linearen Teilraum $\mathfrak{L}'$ des linearen schiefnormierten Raums $\mathfrak{L}$ definiertes, lineares und beschränktes Funktional f kann unter Erhaltung seiner Norm auf ganz $\mathfrak{L}$ ausgedehnt werden.*

Da wir mit $\mathfrak{L}$ auch $\mathfrak{L}^*$ als linearen schiefnormierten Raum erkannt haben, kann man $\mathfrak{L}^{**}$ — den Dualraum des Dualraums oder den Bidualraum — bilden, dessen Punkte die auf $\mathfrak{L}^*$ definierten linearen und beschränkten Funktionale sind. Beispielsweise sind durch $F_{x_0}(f) = f(x_0)$ $(f \in \mathfrak{L}^*)$ derartige Funktionale F_{x_0} für jedes feste $x_0 \in \mathfrak{L}$ definiert, wobei $\|F_{x_0}\| \leqq \|x_0\|$ wegen $f(x_0) \leqq \leqq \|f\| \|x_0\|$. Die Abbildung $x \to F_x$ von $\mathfrak{L}$ in $\mathfrak{L}^{**}$ ist ein Isomorphismus im Sinne der

DEFINITION 6. *Zwei lineare schiefnormierte Räume $\mathfrak{L}_1$ und $\mathfrak{L}_2$ werden isomorph genannt, wenn (mindestens) ein Isomorphismus, d.h. eine lineare und die Norm erhaltende Abbildung T von $\mathfrak{L}_1$ auf $\mathfrak{L}_2$ existiert. Wir schreiben: $\mathfrak{L}_1 \cong \mathfrak{L}_2$ und $\mathfrak{L}_2 = T\mathfrak{L}_1$. Ein Isomorphismus eines Raums auf sich selbst heißt Automorphismus.*

Wir behaupten also:

Satz 3: *Die Abbildung $x \to F_x$ stellt einen Isomorphismus von $\mathfrak{L}$ auf einen linearen Unterraum von $\mathfrak{L}^{**}$ dar.*

BEWEIS. Die Linearität dieser „kanonischen" Abbildung von $\mathfrak{L}$ in $\mathfrak{L}^{**}$ ist trivial. Sei nun $x_0 \neq 0$ ein Punkt aus $\mathfrak{L}$ und $R x_0$ die 0 und x_0 verbindende Gerade. Dann ist durch $f_0(\alpha x_0) = \alpha \|x_0\|$ auf dieser Geraden ein lineares und beschränktes Funktional f_0 definiert, welches nach Satz 2 ohne Änderung seiner Norm 1 auf ganz $\mathfrak{L}$ ausgedehnt werden kann. Aus $F_{x_0}(f_0) = f_0(x_0) = \|x_0\|$ und der schon bewiesenen Ungleichung $\|F_{x_0}\| \leqq \|x_0\|$ folgt $\|F_{x_0}\| = \|x_0\|$, d.h. die kanonische Abbildung $x \to F_x$ erhält die Norm; w.z.b.w.

DEFINITION 7. *Ein linearer schiefnormierter Raum $\mathfrak{L}$ heißt reflexiv, wenn es zu jedem $F \in \mathfrak{L}^{**}$ ein $x \in \mathfrak{L}$ gibt, so daß $F(f) = f(x)$ für alle $f \in \mathfrak{L}^*$ gilt.*

Bei einem reflexiven linearen schiefnormierten Raum $\mathfrak{L}$ ist also nach Satz 3 die kanonische Abbildung $x \to F_x$ ein Isomorphismus von $\mathfrak{L}$ *auf* $\mathfrak{L}^{**}$; daher können $\mathfrak{L}$ und $\mathfrak{L}^{**}$ in diesem Falle mittels der kanonischen Abbildung identifiziert werden. Wir erwähnen, daß z.B. jeder Minkowski-Raum reflexiv ist, da alle linearen Funktionale eines derartigen Raums beschränkt sind und damit Punkte des Dualraums darstellen.

Die Isomorphie von $\mathfrak{L}$ und $\mathfrak{L}^{**}$ im Falle der reflexiven linearen schiefnormierten Räume d.h. in vielen bekannten Fällen läßt die Frage nach der Möglichkeit der Isomorphie von $\mathfrak{L}$ und $\mathfrak{L}^*$ aufkommen. Dies führt zur

DEFINITION 8. *Ein linearer schiefnormierter Raum $\mathfrak{L}$ wird selbstadjungiert genannt, wenn $\mathfrak{L}^*$ und $\mathfrak{L}$ isomorph sind, d.h. wenn (mindestens) ein Isomorphismus A von $\mathfrak{L}^*$ auf $\mathfrak{L}$ existiert. Wir sagen kurz, $\mathfrak{L}$ sei selbstadjungiert mit dem Isomorphismus A (oder $\mathfrak{L}$ besitze A).*

BEMERKUNG. Unsere Definition eines selbstadjungierten *linearen schiefnormierten* Raums ist wohl zu unterscheiden von der ebenfalls möglichen Definition eines selbstadjungierten *linearen topologischen* Raums mit der Forderung der Existenz eines linearen Homöomorphismus zwischen dem Raum und dem aus allen stetigen linearen Funktionalen bestehenden Dualraum (versehen mit der Topologie der gleichmäßigen Konvergenz auf allen beschränkten Teilmengen). Nach der letzteren Definition wären z. B. alle endlichdimensionalen separierten linearen topologischen Räume selbstadjungiert[6]), während — wie wir am Schlusse des nächsten Paragraphen sehen werden — durchaus nicht alle Minkowski-Räume im Sinne unserer, die Normstruktur wesentlich enthaltenden Definition selbstadjungiert sind. Wir bemerken außerdem, daß LORCH eine andere Definition der Selbstadjungiertheit eines Banach-Raums (mit „vollhomogener" Norm) einführt, nach welcher nur der euklidische Raum und der Hilbert-Raum unter den separablen Banach-Räumen selbstadjungiert sein können[7]).

Die Selbstadjungiertheit eines linearen schiefnormierten Raums $\mathfrak{L}$ ist offensichtlich gleichbedeutend mit der Existenz eines *(reellen) inneren Produkts* $\langle x, y\rangle$ für alle $x, y \in \mathfrak{L}$ mit den Eigenschaften: a) $\langle x, y\rangle$ *ist bilinear in* x *und* y, b) $\sup_{\|y\|=1} \langle x, y\rangle = \|x\|$ *und* c) *zu jedem beliebigen* $f \in \mathfrak{L}^*$ *gibt es ein* $x \in \mathfrak{L}$ *mit* $f(y) = \langle x, y\rangle$ *für alle* $y \in \mathfrak{L}$. Hierzu braucht nämlich nur $\langle x, y\rangle = A^{-1}x(y)$ gesetzt zu werden (mit A^{-1} bezeichnen wir den Umkehrisomorphismus von A); und umgekehrt wird durch $f_x(y) = \langle x, y\rangle$ für alle $y \in \mathfrak{L}$ ein Isomorphismus A von $\mathfrak{L}^*$ auf $\mathfrak{L}$ definiert, welcher f_x in x abbildet.

DEFINITION 9. *Ein selbstadjungierter linearer schiefnormierter Raum $\mathfrak{L}$ heißt symmetrisch, wenn (mindestens) ein solcher Isomorphismus A von $\mathfrak{L}^*$ auf $\mathfrak{L}$ existiert, daß für alle x und y aus $\mathfrak{L}$ die Beziehung $A^{-1}x(y) = A^{-1}y(x)$ besteht.*

Diese Definition bedeutet, kurz gesagt, daß ein linearer schiefnormierter Raum dadurch als symmetrisch selbstadjungiert charakterisiert ist, daß man in ihm ein symmetrisches inneres Produkt mit den angegebenen Eigenschaften bilden kann.

Triviale Beispiele von symmetrischen selbstadjungierten und überdies vollständigen, separablen und reflexiven linearen schiefnormierten Räumen sind der Hilbert-Raum und der euklidische Raum im Sinne von

DEFINITION 10. *Ein linearer schiefnormierter Raum heißt (reeller) Hilbert-Raum, wenn er mit dem linearen normierten Raum l_2 aller Folgen $(\xi_1, \xi_2, \ldots)$ mit* $\left(\sum_{i=1}^{\infty} (\xi_i)^2\right)^{\frac{1}{2}} = \|(\xi_1, \xi_2, \ldots)\| < \infty$ *isomorph ist,*

und (entsprechend) von

[6]) Dies folgt aus BOURBAKI [7], I, § 2, No. 3, théorème 2.

[7]) Siehe LORCH [6], S. 468.

DEFINITION 11. *Ein Minkowski-Raum heißt euklidisch, wenn er mit dem gewöhnlichen euklidischen Raum aller Vektoren* $(\xi_1, \ldots, \xi_n)$ *mit der Norm* $\|(\xi_1, \ldots, \xi_n)\| = \left(\sum_{i=1}^{n} (\xi_i)^2\right)^{\frac{1}{2}}$ *isomorph ist*[8]).

Für beide Arten von Räumen kann nämlich durch

$$\langle x, y\rangle = \tfrac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right)$$

ein symmetrisches inneres Produkt mit den erwähnten Eigenschaften definiert werden. (Die Existenz einer Bilinearform $\langle x, y\rangle$ mit $\langle x, x\rangle = \|x\|^2$ auf einem separablen Banach-Raum stellt zusammen mit einer Dimensionsbedingung bekanntlich eine andere mögliche Definition des Hilbert-Raums bzw. des euklidischen Raums dar.)

Als für später sehr nützlich erweist sich schließlich

DEFINITION 12. *Sind* $\mathfrak{L}_1$ *und* $\mathfrak{L}_2$ *zwei lineare schiefnormierte Räume, so wird der Raum aller Punktepaare* $x = (x_1, x_2)$ *mit* $x_1 \in \mathfrak{L}_1$, $x_2 \in \mathfrak{L}_2$, *welcher durch* $\alpha x + \beta y = (\alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2)$ *mit einer linearen sowie durch* $\|x\| = (\|x_1\|^2 + \|x_2\|^2)^{\frac{1}{2}}$ *mit einer Normstruktur versehen ist, das (kartesische) Produkt* $\mathfrak{L}_1 \times \mathfrak{L}_2$ *von* $\mathfrak{L}_1$ *und* $\mathfrak{L}_2$ *genannt.*

Man weist ohne weiteres nach, daß der so definierte Produktraum wiederum linear und schiefnormiert ist. Entsprechend kann auch das Produkt von endlich vielen linearen schiefnormierten Räumen gebildet werden.

BEMERKUNGEN. 1) Wir haben die Punkte des Dualraums $\mathfrak{L}^*$ von $\mathfrak{L}$ als die *beschränkten* und nicht als die (im Sinne der Normtopologie von $|||\ |||$) *stetigen* linearen Funktionale von $\mathfrak{L}$ definiert, wie man dies bei den linearen normierten Räumen tun kann, da dort beide Definitionen äquivalent sind. Der Grund dafür ist die Tatsache, daß zwar *jedes beschränkte lineare Funktional eines linearen schiefnormierten Raums* $\mathfrak{L}$ *stetig ist, aber umgekehrt nicht immer ein stetiges lineares Funktional beschränkt ist.* Man sieht dies folgendermaßen ein: Für jedes beschränkte lineare Funktional f von $\mathfrak{L}$ gilt

$$f(x - x_0) \leqq \|f\|\,\|x - x_0\| \quad \text{und} \quad f(x_0 - x) \leqq \|f\|\,\|x_0 - x\|,$$

also

$$|f(x) - f(x_0)| \leqq \|f\|\,\mathrm{Max}\left(\|x - x_0\|, \|-(x - x_0)\|\right) = \|f\|\,\|\!|x - x_0|\!\|;$$

d.h. f ist im angegebenen Sinne stetig.

Um das Beispiel eines stetigen aber nicht beschränkten linearen Funktionals zu erhalten, betrachten wir den linearen Raum aller Folgen $x = (\xi_1, \xi_2, \ldots)$ mit $\sum_{i=1}^{\infty} |\xi_i| < \infty$ (Addition und skalare Multiplikation erfolge komponentenweise) und erklären in diesem eine Norm durch:

$$\|x\| = \sum_{i=1}^{\infty} \varepsilon_i(x)\,|\xi_i|,$$

wobei $\varepsilon_i(x) = \frac{1}{i}$ falls $\xi_i > 0$ und $\varepsilon_i(x) = 1$ falls $\xi_i \leqq 0$ $(i = 1, 2, \ldots)$.

[8]) Diese Definition ist durch die Tatsache gerechtfertigt, daß die Struktur eines Minkowski-Raums durch seine Norm völlig bestimmt ist.

Man weist leicht nach, daß der so definierte Raum linear und schiefnormiert ist. Jetzt ist das durch $f_0(x) = \sum_{i=1}^{\infty} \xi_i$ gegebene Funktional f_0 linear und wegen $f_0(x-x_0) \leqq \|-(x-x_0)\| \leqq |||x-x_0|||$ sowie $f_0(x_0-x) \leqq |||x-x_0|||$, d.h. $|f_0(x)-f_0(x_0)| \leqq |||x-x_0|||$ stetig. Aber f_0 ist nicht beschränkt, weil für die Punkte $x_j = (\underbrace{0, \ldots, 0}_{j}, j, 0, \ldots)$ $f_0(x_j) = j$ wird $(j = 1, 2, \ldots)$ und $\|x_j\| = 1$ ist.

2) Dieses Beispiel zeigt auch, daß *im allgemeinen die Normtopologie eines linearen schiefnormierten Raums mit dessen linearer Struktur inkompatibel ist.* Bei dem zuletzt betrachteten Raum ist nämlich für die Punkte $y_j = \frac{1}{j} x_j$ $(j = 1, 2, \ldots)$ $\lim_{j \to \infty} \|y_j\| = 0$ aber $\|-y_j\| = 1$; d.h. wie klein auch $\delta > 0$ gewählt wird, stets gilt für die durch $\|x\| < \delta$ gegebene Umgebung U_δ des Nullpunkts: $(-1)\, U_\delta \not\subseteq U_1$. Damit ist aber gezeigt, daß die skalare Multiplikation mit -1 nicht stetig ist, und darum kann die Normtopologie dieses Raums nicht mit seiner linearen Struktur kompatibel sein[9]). Aus diesem Grunde haben wir zur Definition einer Topologie in einem linearen schiefnormierten Raum die Hilfsnorm $|||\ |||$ herangezogen[10]). Wir bemerken noch, daß für einen *endlichdimensionalen* linearen schiefnormierten Raum die Normtopologie mit der linearen Struktur kompatibel ist.

§ 3. Selbstadjungierte schiefe Banach-Räume beliebiger Dimension

Wie wir schon in der Überschrift zu dieser Arbeit zum Ausdruck gebracht haben, werden wir uns nur mit selbstadjungierten schiefen *Banach*-Räumen beschäftigen. Dies bedeutet jedoch gegenüber den selbstadjungierten linearen schiefnormierten Räumen keine Einschränkung, da man zeigen kann:

Satz 4: *Jeder selbstadjungierte lineare schiefnormierte Raum ist ein schiefer Banach-Raum.*

Beweis. Sei $x_1, x_2, \ldots$ eine Fundamentalfolge des selbstadjungierten Raums $\mathfrak{L}$, d.h. zu jedem $\varepsilon > 0$ gebe es ein $N(\varepsilon)$ mit $|||x_i - x_j||| < \varepsilon$ für alle $i, j \geqq N(\varepsilon)$. Diese Folge wird durch den Isomorphismus A^{-1} von $\mathfrak{L}$ auf eine Folge $f_1, f_2, \ldots \in \mathfrak{L}^*$ abgebildet. Bekanntlich ergibt sich nun wegen der Vollständigkeit der reellen Zahlen die Existenz eines linearen Funktionals f_0 mit $f_0(x) = \lim_{i \to \infty} f_i(x)$ für alle $x \in \mathfrak{L}$. $f_0 \in \mathfrak{L}^*$, weil $f_i(x) \leqq |f_i(x)| \leqq |f_i(x) - f_{N(\varepsilon)}(x)| + |f_{N(\varepsilon)}(x)| \leqq (\varepsilon + |||f_{N(\varepsilon)}|||)\, \|x\|$ für alle $i \geqq N(\varepsilon)$ und $x \in \mathfrak{L}$, d.h. $f_0(x) \leqq (\varepsilon + |||f_{N(\varepsilon)}|||)\, \|x\|$ für alle $x \in \mathfrak{L}$. Schließlich gilt: $|f_0(x) - f_i(x)| \leqq \varepsilon \|x\|$ für alle $i \geqq N(\varepsilon)$ und $x \in \mathfrak{L}$, oder anders gesagt: $|||f_0 - f_i||| \leqq \varepsilon$ für alle $i \geqq N(\varepsilon)$. Dies bedeutet aber, daß die f_i gegen f_0 konvergieren, und entsprechend konvergieren die Punkte $x_i = A f_i$ gegen den Punkt $x_0 = A f_0$. Damit ist aber $\mathfrak{L}$ als vollständig erkannt; w.z.b.w.

[9]) Siehe Bourbaki [7], I, § 1, Nr. 1.

[10]) Lorch [8] postuliert offenbar zu diesem Zwecke die Existenz einer Konstanten γ mit $\|-x\| \leqq \gamma \|x\|$ für alle x, wodurch die lineare Struktur des Raumes mit seiner Normtopologie kompatibel wird. Zur Definition der Vollständigkeit des Raums benutzt er allerdings wie wir eine Hilfsnorm.

Von weit größerer Wichtigkeit – insbesondere für den endlichdimensionalen Fall – sind die folgenden Sätze :

Lemma 1: *Wenn T ein Isomorphismus des linearen schiefnormierten Raums $\mathfrak{L}_1$ auf den linearen schiefnormierten Raum $\mathfrak{L}_2$ ist, so stellt der (in der üblichen Weise erklärte) Dualoperator T^* einen Isomorphismus des Dualraums $\mathfrak{L}_2^*$ auf den Dualraum $\mathfrak{L}_1^*$ dar.*

BEWEIS. T ist definiert durch die Beziehung $(T^*g)(x)=g(Tx)$ für alle $x\in\mathfrak{L}_1$ und $g\in\mathfrak{L}_2^*$. Damit wird T^*g ein lineares Funktional auf $\mathfrak{L}_1$, und man errechnet: $\|T^*g\|=\sup_{\|x\|=1} g(Tx)=\sup_{\|Tx\|=1} g(Tx)=\|g\|$; d.h. $T^*g\in\mathfrak{L}_1^*$. T^* ist also ein Isomorphismus von $\mathfrak{L}_2^*$ in $\mathfrak{L}_1^*$. Er bildet sogar $\mathfrak{L}_2^*$ auf $\mathfrak{L}_1^*$ ab, denn jeder beliebige Punkt f von $\mathfrak{L}_1^*$ ist offensichtlich in der Form $f=T^*g$ mit $g\in\mathfrak{L}_2^*$ darstellbar, wobei $g(y)=f(T^{-1}y)$ für alle $y\in\mathfrak{L}_2$ ist; w.z.b.w.

Lemma 1 wird angewandt in

Satz 5: *$\mathfrak{L}_1$ sei selbstadjungiert mit dem Isomorphismus A_1. Außerdem seien $\mathfrak{L}_1$ und $\mathfrak{L}_2=T\mathfrak{L}_1$[11]) isomorph. Dann ist auch L_2 selbstadjungiert und es besitzt den Isomorphismus $A_2=TA_1T^*$.*

Der Beweis dieses Satzes ist eine unmittelbare Folge von Lemma 1.

Satz 6: *Für jeden selbstadjungierten linearen schiefnormierten Raum $\mathfrak{L}$ mit dem Isomorphismus A stellt die Abbildung $C=A(A^*)^{-1}$ einen Isomorphismus von $\mathfrak{L}^{**}$ auf $\mathfrak{L}$ dar. Ist insbesondere $\mathfrak{L}$ außerdem noch reflexiv*[12]*, so ist C ein Automorphismus von $\mathfrak{L}$.*

BEWEIS. Nach Lemma 1 ist $(A^*)^{-1}$ ein Isomorphismus von $\mathfrak{L}^{**}$ auf $\mathfrak{L}^*$, welcher nach Zusammensetzung mit dem Isomorphismus A von $\mathfrak{L}^*$ auf $\mathfrak{L}$ einen Isomorphismus von $\mathfrak{L}^{**}$ auf $\mathfrak{L}$ ergibt. Im Falle eines reflexiven $\mathfrak{L}$ gilt $\mathfrak{L}^{**}=\mathfrak{L}$, d.h. C stellt einen Automorphismus von $\mathfrak{L}$ dar; w.z.b.w.

Bei einem reflexiven selbstadjungierten Raum ist der Automorphismus C von Satz 6 genau dann die identische Abbildung, wenn $A^*=A$ gilt. Daß dadurch gerade die symmetrischen selbstadjungierten Räume gekennzeichnet sind, ist Inhalt von

Satz 7: *Ein selbstadjungierter linearer schiefnormierter Raum mit dem Isomorphismus A ist dann und nur dann (in bezug auf A) symmetrisch, wenn er reflexiv ist und $A^*=A$ gilt.*

BEWEIS. a) Angenommen, der selbstadjungierte Raum $\mathfrak{L}$ sei in bezug auf A symmetrisch, d.h. es existiere ein Isomorphismus A von $\mathfrak{L}^*$ auf $\mathfrak{L}$ mit $A^{-1}x(y)=A^{-1}y(x)$ für alle $x, y\in\mathfrak{L}$. Sei nun F ein beliebiger Punkt von $\mathfrak{L}^{**}$. Wir setzen $y=A(A^*)^{-1}F$ und finden die für alle $f\in\mathfrak{L}^*$ gültige Gleichung:

$$F(f)=(A^*A^{-1}y)(f)=A^{-1}y(Af)=A^{-1}(Af)(y)=f(y);$$

d.h. $\mathfrak{L}$ ist reflexiv mit der kanonischen Abbildung $y\to F_y=A^*A^{-1}y$. Identifiziert man nun y mit F_y, so folgt in der Tat $A^*=A$.

---

[11]) Wegen dieser Schreibweise siehe Definition 6.

[12]) Aus $\mathfrak{L}^{**}\cong\mathfrak{L}$ selbst folgt noch nicht notwendig die Reflexivität von $\mathfrak{L}$, d.h. die Tatsache, daß auch die kanonische Abbildung $x\to F_x$ ein Isomorphismus von $\mathfrak{L}$ auf $\mathfrak{L}^{**}$ ist.

b) Wir setzen jetzt $\mathfrak{L}$ als reflexiv voraus und nehmen $A^* = A$ an. Dann gilt für alle $x, y \in \mathfrak{L}$:

$$A^{-1}x(y) = A^{-1}x(AA^{-1}y) = A^*A^{-1}x(A^{-1}y) = A^{-1}y(x),$$

d.h. $\mathfrak{L}$ ist in bezug auf A symmetrisch; w.z.b.w.

Satz 8: *Wenn die selbstadjungierten linearen schiefnormierten Räume $\mathfrak{L}_1$ bzw. $\mathfrak{L}_2$ die Isomorphismen A_1 bzw. A_2 besitzen, ist ihr Produkt $\mathfrak{L}_1 \times \mathfrak{L}_2$ selbstadjungiert mit dem Isomorphismus $A = A_1 \times A_2$, welcher durch $Af = (A_1 f_1, A_2 f_2)$ $\left(f \in (\mathfrak{L}_1 \times \mathfrak{L}_2)^*, f_1 \in \mathfrak{L}_1^*, f_2 \in \mathfrak{L}_2^*\right)$ erklärt ist, wobei sich f_1 bzw. f_2 aus f durch $f_1(x_1) = f(x_1, 0)$ bzw. $f_2(x_2) = f(0, x_2)$ für alle $x_1 \in \mathfrak{L}_1$, $x_2 \in \mathfrak{L}_2$ herleitet.*

Dem Beweis dieses Satzes schicken wir voraus:

Lemma 2: *Für zwei beliebige lineare schiefnormierte Räume $\mathfrak{L}_1$ und $\mathfrak{L}_2$ gilt: $(\mathfrak{L}_1 \times \mathfrak{L}_2)^* \cong \mathfrak{L}_1^* \times \mathfrak{L}_2^*$. Die Isomorphie wird hierbei durch einen Isomorphismus bewirkt, welcher jedem Punkt $f \in (\mathfrak{L}_1 \times \mathfrak{L}_2)^*$ denjenigen Punkt $(f_1, f_2) \in \mathfrak{L}_1^* \times \mathfrak{L}_2^*$ $(f_i \in \mathfrak{L}_i^*)$ zuordnet, der durch die Beziehungen $f_1(x_1) = f(x_1, 0)$ und $f_2(x_2) = f(0, x_2)$ für alle $x_i \in \mathfrak{L}_i$ $(i = 1, 2)$ eindeutig bestimmt ist.*

Beweis. Durch $f_1(x_1) = f(x_1, 0)$ und $f_2(x_2) = f(0, x_2)$ $(x_i \in \mathfrak{L}_i)$ wird offensichtlich jedem $f \in (\mathfrak{L}_1 \times \mathfrak{L}_2)^*$ eindeutig ein $f_i \in \mathfrak{L}_i$ zugeordnet $(i = 1, 2)$. Damit ist durch $f \to (f_1, f_2)$ eine lineare Abbildung von $(\mathfrak{L}_1 \times \mathfrak{L}_2)^*$ in $\mathfrak{L}_1^* \times \mathfrak{L}_2^*$ erklärt. Sie ist isometrisch, denn es gilt $\|(f_1, f_2)\| = \|f\|$ wegen:

$$\|(f_1, f_2)\| = (\|f_1\|^2 + \|f_2\|^2)^{\frac{1}{2}} = \sup_{\|x_1\| = \frac{\|f_1\|}{\|(f_1, f_2)\|}} f_1(x_1) + \sup_{\|x_2\| = \frac{\|f_2\|}{\|(f_1, f_2)\|}} f_2(x_2) \leqq$$

$$\leqq \sup_{\|(x_1, x_2)\| = 1} \left(f_1(x_1) + f_2(x_2)\right) = \sup_{\|(x_1, x_2)\| = 1} f(x_1, x_2) = \|f\| \leqq$$

$$\leqq \sup_{\|(x_1, x_2)\| = 1} \left(\|f_1\| \, \|x_1\| + \|f_2\| \, \|x_2\|\right) \leqq$$

$$\leqq \sup_{\|x_1\|^2 + \|x_2\|^2 = 1} (\|f_1\|^2 + \|f_2\|^2)^{\frac{1}{2}} (\|x_1\|^2 + \|x_2\|^2)^{\frac{1}{2}} = \|(f_1, f_2)\|.$$

Schließlich ist jeder beliebige Punkt $(f_1, f_2) \in \mathfrak{L}_1^* \times \mathfrak{L}_2^*$ Bild des durch $f(x_1, x_2) = f_1(x_1) + f_2(x_2)$ definierten Punktes $f \in (\mathfrak{L}_1 \times \mathfrak{L}_2)^*$, d.h. die betrachtete Abbildung bewirkt einen Isomorphismus von $(\mathfrak{L}_1 \times \mathfrak{L}_2)^*$ auf $\mathfrak{L}_1^* \times \mathfrak{L}_2^*$; w.z.b.w.

Der Beweis von Satz 8 folgt jetzt unmittelbar aus Lemma 2.

Durch Anwendung von Lemma 2 gewinnt man das von Lorch[13]) angegebene Beispiel eines nichttrivialen unendlichdimensionalen, symmetrischen selbstadjungierten linearen normierten Raums $\mathfrak{B}$: Sei l_p der Raum aller Folgen $(\xi_1, \xi_2, \ldots)$ mit $\|(\xi_1, \xi_2, \ldots)\| = \left(\sum_{i=1}^{\infty} |\xi_i|^p\right)^{1/p} < \infty$ $(1 < p < 2)$. Dann werde $\mathfrak{B} = l_p \times l_q$ gesetzt $\left(\frac{1}{p} + \frac{1}{q} = 1\right)$, und man findet wegen Lemma 2 und wegen $l_p^* \cong l_q$, $l_q^* \cong l_p$: $(l_p \times l_q)^* \cong l_q \times l_p$. $l_q \times l_p$ läßt sich aber mittels $(x_2, x_1) \to (x_1, x_2)$ isomorph auf $l_p \times l_q$ abbilden $(x_1 \in l_p, x_2 \in l_q)$; d.h. $\mathfrak{B}$ ist selbstadjungiert und, wie man leicht sieht, symmetrisch. Andererseits ist $\mathfrak{B}$ sicher kein Hilbert-Raum, da $\mathfrak{B}$ den zweidimensionalen linearen Unterraum $l_p^{(2)}$ aller Punkte

[13]) Vgl. [6], S. 473.

$x=(x_1, 0)$ mit $x_1=(\xi_1, \xi_2, 0, 0, \ldots)$ und $\|x\|=(|\xi_1|^p+|\xi_2|^p)^{1/p}$ enthält, welcher unter der Annahme, $\mathfrak{B}$ sei ein Hilbert-Raum, euklidisch sein müßte. Dann wäre aber $\|x\|$ gleich der Quadratwurzel einer positiv definiten quadratischen Form in den Variablen ξ_1 und ξ_2, was zu einem Widerspruch führt. Dieses Beispiel läßt sich übrigens dahingehend verallgemeinern, daß schon der Raum $\mathfrak{L}\times\mathfrak{L}^*$ nichttrivial und symmetrisch selbstadjungiert ist, wenn man nur $\mathfrak{L}$ als einen reflexiven linearen schiefnormierten Raum voraussetzt, der weder ein Hilbert-Raum noch euklidisch ist.

Im Gegensatz hierzu wollen wir zum Schluß noch zwei Beispiele von nicht selbstadjungierten Banach-Räumen angeben. So sind einmal alle Räume l_p mit $1<p<\infty$ und $p\neq 2$ nicht selbstadjungiert, da aus $l_p\cong l_q$ mit $\frac{1}{p}+\frac{1}{q}=1$ $p=q$ d.h. $p=2$ folgt[14]) (dasselbe gilt auch für die entsprechenden Funktionenräume L_p mit $1<p<\infty$ und $p\neq 2$[15])). Es gibt aber auch nicht selbstadjungierte *Minkowski-Räume*, wie etwa den dreidimensionalen Vektorraum $l_1^{(3)}$, in dem durch $\|x\|=|\xi_1|+|\xi_2|+|\xi_3|$ eine (vollhomogene) Norm definiert ist. Der Dualraum $l_\infty^{(3)}$ von $l_1^{(3)}$ besitzt nämlich offensichtlich die Norm $\|f\|=$ $=\mathrm{Max}(|\alpha_1|, |\alpha_2|, |\alpha_3|)$. Wäre nun $l_1^{(3)}$ selbstadjungiert, d.h. wäre $l_1^{(3)}\cong l_\infty^{(3)}$, so müßte der Würfel $\|f\|=1$ bei Anwendung der linearen Abbildung A in das Oktaeder $\|x\|=1$ übergehen, eine Unmöglichkeit. Daraus folgt in der Tat, daß $l_1^{(3)}$ (wie auch $l_\infty^{(3)}$) nicht selbstadjungiert sein kann.

§ 4. Selbstadjungierte Minkowski-Räume und ihr kanonischer Isomorphismus

Von jetzt ab machen wir die Einschränkung, daß ein selbstadjungierter linearer schiefnormierter Raum immer von endlicher Dimension n, d.h. ein selbstadjungierter Minkowski-Raum $\mathfrak{M}$ sei. Dann läßt sich ein Punkt $x\in\mathfrak{M}$ und ein Punkt $f\in\mathfrak{M}^*$ in der Form $x=(\xi_1, \ldots, \xi_n)$ und $f=(\alpha_1, \ldots, \alpha_n)$ darstellen, wenn f durch die Gleichung $f(x)=\alpha_1\xi_1+\cdots+\alpha_n\xi_n$ gegeben ist; und die Beziehung $x=Af$ schreibt sich „in Koordinaten": $\xi_i=\sum_{k=1}^{n}\alpha_{ik}\alpha_k$ $(i=1, 2, \ldots, n)$. Hierbei ist wie immer A ein Isomorphismus von $\mathfrak{M}^*$ auf $\mathfrak{M}$, welcher also in eindeutiger Weise durch die quadratische Matrix $\mathfrak{A}$ mit den Elementen α_{ik} bestimmt ist. Allgemeiner ist jeder Isomorphismus T eines Minkowski-Raums $\mathfrak{M}_1$ auf einen ebensolchen $\mathfrak{M}_2$ durch eine quadratische Matrix $\mathfrak{T}$ eindeutig bestimmt. Wir drücken diesen Sachverhalt einfach so aus, daß wir sagen, $\mathfrak{A}$ *bzw.* $\mathfrak{T}$ *sei die Matrix von* A *bzw. von* T. Man sieht nun sofort, daß der duale Isomorphismus A^* von $\mathfrak{M}^*$ auf $\mathfrak{M}^{**}=\mathfrak{M}$ die transponierte Matrix $\mathfrak{A}^*$ von $\mathfrak{A}$ mit den Elementen $\alpha_{ik}^*=\alpha_{ki}$ besitzt; setzen wir nämlich $g=(\beta_1, \ldots, \beta_n)$ und $y=A^*g=(\eta_1, \ldots, \eta_n)$, so folgt aus der für alle $f\in\mathfrak{M}^*$ gültigen Beziehung $f(y)=(A^*g)(f)=g(Af)$ die Koordinatendarstellung:

$$\sum_{i=1}^{n}\alpha_i\eta_i=\sum_{i=1}^{n}\beta_i\sum_{k=1}^{n}\alpha_{ik}\alpha_k=\sum_{i=1}^{n}\alpha_i\sum_{k=1}^{n}\alpha_{ki}\beta_k \quad \text{d.h.} \quad \eta_i=\sum_{k=1}^{n}\alpha_{ki}\beta_k \ (i=1, 2, \ldots, n).$$

[14]) Siehe [9], S. 205, théorème 7.
[15]) Siehe [9], S. 203, théorème 4.

Analog ist $\mathfrak{T}^*$ die Matrix von T^*; und es erweist sich ferner $\mathfrak{C} = \mathfrak{A}(\mathfrak{A}^*)^{-1}$ als die Matrix von $C = A(A^*)^{-1}$ (vgl. Satz 6) sowie $\mathfrak{T}\mathfrak{A}\mathfrak{T}^*$ als die Matrix von TAT^* (vgl. Satz 5) und schließlich $\left(\begin{array}{c|c} \mathfrak{A}_1 & 0 \\ \hline 0 & \mathfrak{A}_2 \end{array}\right) = \mathfrak{A}_1 \dot{+} \mathfrak{A}_2$[16]) als die Matrix von $A_1 \times A_2$ (vgl. Satz 8).

Um einen für die folgenden Überlegungen fundamentalen Satz formulieren zu können, benötigen wir nun:

DEFINITION 13. *Zwei (quadratische) Matrizen $\mathfrak{A}_1$ und $\mathfrak{A}_2$ werden kongruent bzw. ähnlich genannt, wenn eine solche nichtsinguläre (quadratische) Matrix $\mathfrak{T}$ existiert, daß $\mathfrak{A}_2 = \mathfrak{T}\mathfrak{A}_1\mathfrak{T}^*$ bzw. $\mathfrak{A}_2 = \mathfrak{T}\mathfrak{A}_1\mathfrak{T}^{-1}$ gilt.*

DEFINITION 14. *Eine (quadratische) Matrix $\mathfrak{A}$ heißt normal, wenn $\mathfrak{A}\mathfrak{A}^* = \mathfrak{A}^*\mathfrak{A}$ ist.*

Dann ist unser erstes Hauptresultat:

Satz 9: *Wenn $\mathfrak{M}$ ein selbstadjungierter Minkowski-Raum mit dem Isomorphismus A ist, so ist die Matrix $\mathfrak{A}$ von A mit einer normalen Matrix $\mathfrak{A}^{(1)}$ kongruent.*

BEWEIS. Die Gruppe G der Matrizen aller Automorphismen von $\mathfrak{M}$ enthält nach Satz 6 das Element $\mathfrak{C} = \mathfrak{A}(\mathfrak{A}^*)^{-1}$. Da die Automorphismen von $\mathfrak{M}$ alle die (in gewöhnlichem Sinne) beschränkte Punktmenge $\|x\| \leqq 1$ in sich überführen, sind alle Elemente aller Matrizen von G ihrem Absolutbetrag nach gleichmäßig beschränkt. Nun folgt daraus durch Anwendung eines Satzes von J. v. NEUMANN[17]), daß die Matrix $\mathfrak{C}$ einer orthogonalen Matrix $\mathfrak{C}^{(1)}$ ähnlich ist. Es gibt also eine nichtsinguläre Matrix $\mathfrak{T}$ mit $\mathfrak{C}^{(1)} = \mathfrak{T}\mathfrak{C}\mathfrak{T}^{-1} = (\mathfrak{C}^{(1)*})^{-1}$. Jetzt setzen wir $\mathfrak{A}^{(1)} = \mathfrak{T}\mathfrak{A}\mathfrak{T}^*$ und finden:

$$\mathfrak{A}^{(1)}(\mathfrak{A}^{(1)*})^{-1} = (\mathfrak{T}\mathfrak{A}\mathfrak{T}^*)(\mathfrak{T}^{*-1}\mathfrak{A}^{*-1}\mathfrak{T}^{-1}) = \mathfrak{T}\mathfrak{C}\mathfrak{T}^{-1} = \mathfrak{C}^{(1)} = (\mathfrak{C}^{(1)*})^{-1} = (\mathfrak{A}^{(1)*})^{-1}\mathfrak{A}^{(1)}$$

oder

$$\mathfrak{A}^{(1)*}\mathfrak{A}^{(1)} = \mathfrak{A}^{(1)}\mathfrak{A}^{(1)*}.$$

$\mathfrak{A}^{(1)}$ ist also in der Tat normal und mit $\mathfrak{A}$ kongruent; w. z. b. w.

Für die normalen Matrizen kann man nun beweisen:

Lemma 3: *Jede n-reihige, normale und nichtsinguläre Matrix $\mathfrak{A}^{(1)}$ ist mit einer Matrix $\mathfrak{A}^{(2)}$ der Gestalt* $\mathfrak{A}^{(2)} = \underbrace{1 \dot{+} \cdots \dot{+} 1}_{l} \dot{+} \mathfrak{d}_1 \dot{+} \cdots \dot{+} \mathfrak{d}_m \dot{+} \underbrace{(-1) \dot{+} \cdots \dot{+} (-1)}_{n-2m-l}$

kongruent[16]), *wobei* $\mathfrak{d}_j = \begin{pmatrix} \cos\varphi_j & \sin\varphi_j \\ -\sin\varphi_j & \cos\varphi_j \end{pmatrix}$ $(j = 1, 2, \ldots, m)$ *und*

$$0 < \varphi_1 \leqq \varphi_2 \leqq \cdots \leqq \varphi_m < \pi \text{ ist.}$$

BEWEIS. Man weiß, daß jede normale Matrix bezüglich einer unitären Matrix einer Matrix in Diagonalform ähnlich ist, deren Diagonalelemente genau die reellen und paarweise konjugiert komplexen Eigenwerte der (reellen) Ausgangsmatrix sind[18]). Nun besitzt $\mathfrak{A}^{(1)}$ als nichtsinguläre Matrix im ganzen n,

[16]) Zu dieser Bezeichnungsweise vgl. H. BOERNER: Darstellungen von Gruppen, S. 16. Berlin: Springer 1955.

[17]) Siehe [*10*], theorem 19.

[18]) Einen Beweis dieser Tatsache findet man bei W. SCHMEIDLER [*11*], S. 57, Satz 26.

alles von 0 verschiedene Eigenwerte, nämlich etwa $\varrho_1, \ldots, \varrho_l; \sigma_1 e^{i\varphi_1}, \sigma_1 e^{-i\varphi_1}, \ldots, \sigma_m e^{i\varphi_m}, \sigma_m e^{-i\varphi_m}; -\tau_1, \ldots, -\tau_{n-2m-l}$ ($\varrho_i, \sigma_j, \tau_k$ reell und >0; $0<\varphi_1 \leqq \cdots \leqq \varphi_m < \pi$), so daß die eben angeführte Tatsache, auf $\mathfrak{A}^{(1)}$ angewandt und ins Reelle übersetzt, bedeutet: Es existiert eine orthogonale Matrix $\mathfrak{O}$, so daß

$$\mathfrak{A}_1^{(1)} = \varrho_1 \dotplus \cdots \dotplus \varrho_l \dotplus \sigma_1 \mathfrak{d}_1 \dotplus \cdots \dotplus \sigma_m \mathfrak{d}_m \dotplus (-\tau_1) \dotplus \cdots \dotplus (-\tau_{n-2m-l}) = \mathfrak{O}\,\mathfrak{A}^{(1)}\,\mathfrak{O}^{-1} = \mathfrak{O}\,\mathfrak{A}^{(1)}\,\mathfrak{O}^*$$

mit $\mathfrak{d}_j = \begin{pmatrix} \cos\varphi_j & \sin\varphi_j \\ -\sin\varphi_j & \cos\varphi_j \end{pmatrix}$ $(j = 1, 2, \ldots, m)$ gilt[19]). Weiter sieht man:

$$\mathfrak{A}_1^{(1)} = \mathfrak{T}\,\mathfrak{A}^{(2)}\,\mathfrak{T}^*,$$

wobei

$$\mathfrak{T} = (\varrho_1)^{\frac{1}{2}} \dotplus \cdots \dotplus (\varrho_l)^{\frac{1}{2}} \dotplus (\sigma_1)^{\frac{1}{2}} \dotplus (\sigma_1)^{\frac{1}{2}} \dotplus \cdots \dotplus (\sigma_m)^{\frac{1}{2}} \dotplus (\sigma_m)^{\frac{1}{2}} \dotplus (\tau_1)^{\frac{1}{2}} \dotplus \cdots \dotplus (\tau_{n-2m-l})^{\frac{1}{2}}$$

und $\mathfrak{A}^{(2)}$ von der in Lemma 3 angegebenen Gestalt ist. Aus $\mathfrak{A}^{(2)} = (\mathfrak{T}^{-1}\mathfrak{O})\,\mathfrak{A}^{(1)}\,(\mathfrak{T}^{-1}\mathfrak{O})^*$ folgt aber die behauptete Kongruenz von $\mathfrak{A}^{(1)}$ und $\mathfrak{A}^{(2)}$; w.z.b.w.

Lemma 4: *Wenn die Matrix $\mathfrak{A}_1^{(2)}$ in der in Lemma 3 vorkommenden, speziellen Gestalt die m_1 Winkel $(\varphi_1)_j$ besitzt und mit einer ebensolchen, die m_2 Winkel $(\varphi_2)_{j'}$ besitzenden Matrix $\mathfrak{A}_2^{(2)}$ kongruent ist, so ist $m_1 = m_2 = m$, und es gilt entweder $(\varphi_1)_j = (\varphi_2)_{j'}$ oder $(\varphi_1)_j = \pi - (\varphi_2)_{j'}$ $(1 \leqq j, j' \leqq m)$. Außerdem sind die Ränge und die Trägheitsindizes der symmetrisierten Matrizen $\frac{1}{2}(\mathfrak{A}_1^{(2)} + \mathfrak{A}_1^{(2)*})$ und $\frac{1}{2}(\mathfrak{A}_2^{(2)} + \mathfrak{A}_2^{(2)*})$ gleich.*

BEWEIS. Da $\mathfrak{A}_1^{(2)}$ und $\mathfrak{A}_2^{(2)}$ kongruente Matrizen sind, gibt es eine nichtsinguläre Matrix $\mathfrak{T}$, so daß $\mathfrak{A}_2^{(2)} = \mathfrak{T}\,\mathfrak{A}_1^{(2)}\,\mathfrak{T}^*$ und damit

$$\mathfrak{A}_2^{(2)}(\mathfrak{A}_2^{(2)*})^{-1} = \mathfrak{T}\left(\mathfrak{A}_1^{(2)}(\mathfrak{A}_1^{(2)*})^{-1}\right)\mathfrak{T}^{-1}$$

ist. Aber $\mathfrak{A}_1^{(2)}$ und $\mathfrak{A}_2^{(2)}$ sind orthogonale Matrizen, d.h. es gilt $\mathfrak{A}_1^{(2)*} = (\mathfrak{A}_1^{(2)})^{-1}$ und $\mathfrak{A}_2^{(2)*} = (\mathfrak{A}_2^{(2)})^{-1}$. Also sind die Matrizen $(\mathfrak{A}_1^{(2)})^2$ und $(\mathfrak{A}_2^{(2)})^2 = \mathfrak{T}(\mathfrak{A}_1^{(2)})^2\mathfrak{T}^{-1}$ ähnlich. Daraus ergibt sich, daß die Menge aller Eigenwerte von $(\mathfrak{A}_1^{(2)})^2$, nämlich

$$\left\{\underbrace{1, \ldots, 1}_{n-2m_1}, e^{2(\varphi_1)_1 i}, e^{-2(\varphi_1)_1 i}, \ldots, e^{2(\varphi_1)_{m_1} i}, e^{-2(\varphi_1)_{m_1} i}\right\},$$

mit der Menge

$$\left\{\underbrace{1, \ldots, 1}_{n-2m_2}, e^{2(\varphi_2)_1 i}, e^{-2(\varphi_2)_1 i}, \ldots, e^{2(\varphi_2)_{m_2} i}, e^{-2(\varphi_2)_{m_2} i}\right\}$$

aller Eigenwerte von $(\mathfrak{A}_2^{(2)})^2$ als Ganzes übereinstimmt. Deshalb ist $m_1 = m_2 = m$, sowie entweder $(\varphi_1)_j = (\varphi_2)_{j'}$ oder $(\varphi_1)_j = \pi - (\varphi_2)_{j'}$ $(1 \leqq j, j' \leqq m)$. Außerdem folgt aus $\mathfrak{A}_2^{(2)} = \mathfrak{T}\,\mathfrak{A}_1^{(2)}\,\mathfrak{T}^*$ die Beziehung $\frac{1}{2}(\mathfrak{A}_2^{(2)} + \mathfrak{A}_2^{(2)*}) = \mathfrak{T}\left[\frac{1}{2}(\mathfrak{A}_1^{(2)} + \mathfrak{A}_1^{(2)*})\right]\mathfrak{T}^*$, so daß nach dem Sylvesterschen Satz Ränge und Trägheitsindizes der symmetrisierten Matrizen $\frac{1}{2}(\mathfrak{A}_1^{(2)} + \mathfrak{A}_1^{(2)*})$ und $\frac{1}{2}(\mathfrak{A}_2^{(2)} + \mathfrak{A}_2^{(2)*})$ gleich sind. Dies bedeutet übrigens genau, daß die Anzahlen der von 0 verschiedenen sowie die Anzahlen der positiven Elemente der Hauptdiagonale von $\mathfrak{A}_1^{(2)}$ und von $\mathfrak{A}_2^{(2)}$ jeweils dieselben sind; w.z.b.w.

[19]) Man sieht dies genauso ein wie im Spezialfall $\varrho_i = \sigma_j = \tau_k = 1$ bei SCHREIER-SPERNER [12], S. 94–98.

BEMERKUNG. Auf Grund von Satz 5, Satz 9, Lemma 3 und Lemma 4 kann man die aus der Matrix $\mathfrak{A}$ des Isomorphismus A des selbstadjungierten Minkowski-Raums $\mathfrak{M}$ gewonnenen Größen m, $\operatorname{Min}(\varphi_j, \pi - \varphi_j)$ $(j = 1, 2, \ldots, m)$ sowie den Rang und den Trägheitsindex von $\frac{1}{2}(\mathfrak{A} + \mathfrak{A}^*)$ als *Invarianten von A* bezüglich beliebiger Isomorphismen von $\mathfrak{M}$ ansehen.

Außerdem benötigen wir noch den folgenden Satz von KRONECKER aus der Theorie der diophantischen Approximationen[20]:

Lemma 5: *Wenn die reellen Zahlen $\delta_1, \ldots, \delta_p, 1$ über dem Körper der rationalen Zahlen linear unabhängig sind, dann liegen die im p-dimensionalen, affinen Raum gelegenen Punkte $P_\nu = (\varepsilon_1^{(\nu)}, \ldots, \varepsilon_p^{(\nu)})$ mit den Koordinaten $\varepsilon_r^{(\nu)} = \nu\delta_r - [\nu\delta_r]$ $(\nu = 0, 1, 2, \ldots; 1 \leqq r \leqq p)$ in dem p-dimensionalen Einheitswürfel dieses Raums dicht.*

Natürlich können mehrere Isomorphismen die Isomorphie eines selbstadjungierten Minkowski-Raums $\mathfrak{M}$ mit seinem Dualraum bewirken, und es ist klar, daß wenn $\mathfrak{A}$ die Matrix eines derartigen Isomorphismus A ist, die Menge der Matrizen aller Isomorphismen von $\mathfrak{M}^*$ auf $\mathfrak{M}$ durch $G\mathfrak{A}$ gegeben wird, wobei G die Gruppe der Matrizen aller Automorphismen von $\mathfrak{M}$ darstellt. Der folgende Satz sagt nun aus, daß mindestens eine der Matrizen aus $G\mathfrak{A}$ von einer sehr einfachen Gestalt ist, deren zugehörigen Isomorphismus wir als einen *„kanonischen" Isomorphismus von* $\mathfrak{M}$ bezeichnen wollen.

Satz 10: *Ein selbstadjungierter Minkowski-Raum $\mathfrak{M}$ besitzt immer (mindestens) einen kanonischen Isomorphismus $A^{(0)}$ mit der Matrix $\mathfrak{A}^{(0)}$ der Gestalt* $\mathfrak{A}^{(0)} = \mathfrak{T}\big(\underbrace{1 \dotplus \cdots \dotplus 1}_{n-2m-l} \dotplus \mathfrak{d}_1 \dotplus \cdots \dotplus \mathfrak{d}_m \dotplus \underbrace{(-1) \dotplus \cdots \dotplus (-1)}_{l}\big)\mathfrak{T}^*$, *wobei $\mathfrak{T}$ eine nichtsinguläre Matrix darstellt und* $\mathfrak{d}_j = \begin{pmatrix} \cos\varphi_j & \sin\varphi_j \\ -\sin\varphi_j & \cos\varphi_j \end{pmatrix}$ *mit* $\varphi_j = \frac{l_j}{2^{\lambda_j}}\pi$ *ist $(j = 1, 2, \ldots, m)$. Hierin ist jeweils λ_j eine ganze und l_j eine ungerade ganze Zahl, und es gelten die Ungleichungen* $0 < \varphi_1 \leqq \varphi_2 \leqq \cdots \leqq \varphi_m < \pi$ $(j = 1, 2, \ldots, m)$.

BEWEIS. Wir gehen aus von der Matrix $\mathfrak{A}$ irgendeines Isomorphismus A von $\mathfrak{M}^*$ auf $\mathfrak{M}$. Nach Satz 9 ist $\mathfrak{A}$ mit einer normalen Matrix $\mathfrak{A}^{(1)}$ kongruent. Da $\mathfrak{A}$ und damit auch $\mathfrak{A}^{(1)}$ n-reihig und nichtsingulär ist, kann man Lemma 3 anwenden und findet für die dort auftretende Matrix $\mathfrak{A}^{(2)}$ die Beziehung $\mathfrak{A}^{(2)} = \mathfrak{T}_2\mathfrak{A}\mathfrak{T}_2^*$, wobei $\mathfrak{T}_2$ eine nichtsinguläre Matrix darstellt. $\mathfrak{A}$ hat also schon beinahe die in Satz 10 geforderte Form, nur wissen wir natürlich nicht, ob die Winkel φ_j nur die darin vorgeschriebenen speziellen Werte annehmen. Um dies zu erreichen, gehen wir in mehreren Schritten vor.

1. $\mathfrak{T}_2$ ist Matrix einer linearen Abbildung T_2 des n-dimensionalen, affinen Raums auf sich. Definiert man in dem letzteren durch $\|x\| = \|(T_2)^{-1}x\|_{\mathfrak{M}}$[21] eine Norm, so erhält man einen Minkowski-Raum $\mathfrak{M}^{(2)}$. T_2 bewirkt jetzt offenbar einen Isomorphismus von $\mathfrak{M}$ auf $\mathfrak{M}^{(2)} = T_2\mathfrak{M}$; und nach Satz 5 ist deshalb $\mathfrak{M}^{(2)}$ selbstadjungiert mit dem Isomorphismus $A^{(2)}$, dessen Matrix gerade $\mathfrak{A}^{(2)}$ ist. Wir behaupten nun als erstes:

---

[20]) Siehe [13], S. 83, Satz 5.

[21]) Hierin bedeutet $\|\ \|_{\mathfrak{M}}$ die Norm des Minkowski-Raums $\mathfrak{M}$.

$\mathfrak{M}^{(2)}$ besitzt auch einen Isomorphismus $A^{(3)}$ mit der Matrix $\mathfrak{A}^{(3)}$, welche dieselbe Gestalt wie $\mathfrak{A}^{(2)}$ mit Ausnahme der in Lemma 3 angegebenen Winkelungleichungen besitzt und bei der außerdem alle Größen $\delta_j = \frac{\varphi_j}{\pi}$ rationale Zahlen sind $(0 \leqq \delta_j < 2;\ 1 \leqq j \leqq m)$.

Dies kann man folgendermaßen einsehen: Entweder haben schon alle δ_j von $\mathfrak{A}^{(2)}$ diese Eigenschaft und man hat nichts zu beweisen, oder aber dies ist nicht so. Im letzteren Falle können wir annehmen, daß die δ_j für alle $j < j_0$ rational und für $j = j_0$ irrational sind $(1 \leqq j_0 \leqq m)$. Jetzt ist $\mathfrak{A}^{(2)} (\mathfrak{A}^{(2)*})^{-1} = (\mathfrak{A}^{(2)})^2$ die Matrix eines Automorphismus von $\mathfrak{M}^{(2)}$, und somit bildet die Menge aller (ganzzahligen) Potenzen dieser Matrix eine zyklische Untergruppe G_2' der Gruppe G_2 der Matrizen aller Automorphismen von $\mathfrak{M}^{(2)}$. Die zweireihigen Hauptminoren der Elemente von G_2' mit dem Index j sind die Rotationsmatrizen $(\mathfrak{d}_j)^{2\nu}$ mit ganzzahligem ν und den zugehörigen Drehwinkeln $(\nu\delta_j - [\nu\delta_j])\,2\pi$ $(1 \leqq j \leqq m)$. Indem man nun Lemma 5 für die über dem Körper der rationalen Zahlen linear unabhängigen Zahlen δ_{j_0} und 1 anwendet und beachtet, daß G_2 eine (im Sinne der natürlichen Topologie des Raums aller n-reihigen, quadratischen Matrizen) kompakte Menge darstellt, findet man in der (in G_2 liegenden) abgeschlossenen Hülle von G_2' eine Matrix $\mathfrak{G}_2$, welche die direkte Summe ihrer einreihigen Hauptminoren (1) und ihrer zweireihigen Hauptminoren $\begin{pmatrix} \cos\psi_j & \sin\psi_j \\ -\sin\psi_j & \cos\psi_j \end{pmatrix}$ $(1 \leqq j \leqq m)$ ist, wobei $\psi_{j_0} = \delta_{j_0}\pi$ gilt. Außerdem sind die Größen $\varepsilon_j = \frac{\psi_j}{\pi}$ für $j < j_0$ rational, weil aus der Tatsache, daß die δ_j für $j < j_0$ rational sind, folgt, daß für diese Indizes j die von $(\mathfrak{d}_j)^2$ erzeugte zyklische Gruppe jeweils eine endliche Ordnung besitzt. Jetzt ist $(\mathfrak{G}_2)^{-1}\mathfrak{A}^{(2)}$ Matrix eines Isomorphismus von $\mathfrak{M}^{(2)*}$ auf $\mathfrak{M}^{(2)}$, welche (mit Ausnahme der Winkelungleichungen) dieselbe Gestalt wie $\mathfrak{A}^{(2)}$ und die Winkel $\varphi_j' = \delta_j'\pi$ besitzt $(0 \leqq \delta_j' < 2)$. Hierbei gilt: $\delta_j' \equiv \delta_j - \varepsilon_j \pmod 2$ $(1 \leqq j \leqq m)$ und speziell $\delta_{j_0}' = 0$; d. h. alle δ_j' mit $j \leqq j_0$ sind rational. Auf dieselbe Weise fortfahrend gelangt man nach endlich vielen Schritten schließlich zu einem Isomorphismus $A^{(3)}$ von $\mathfrak{M}^{(2)*}$ auf $\mathfrak{M}^{(2)}$, dessen Matrix $\mathfrak{A}^{(3)}$ die in unserer ersten Teilbehauptung geforderten Eigenschaften hat.

2. Die eben bewiesene Behauptung von 1. läßt sich noch verschärfen. Es gilt nämlich:

$\mathfrak{M}^{(2)}$ besitzt auch einen Isomorphismus $A^{(4)}$ mit der Matrix $\mathfrak{A}^{(4)}$, welche dieselbe Gestalt wie $\mathfrak{A}^{(2)}$ mit Ausnahme der in Lemma 3 angegebenen Winkelungleichungen besitzt und für die entweder $\delta_j = 0$ oder $\delta_j = \frac{l_j}{2^{\lambda_j}}$ mit ganzem λ_j und ungeradem ganzen l_j gilt, wobei l_j noch der Einschränkung $0 < l_j < 2^{\lambda_j+1}$ unterworfen ist $(1 \leqq j \leqq m)$.

Um dies einzusehen, stellen wir die nach 1. rationalen Größen δ_j von $\mathfrak{A}^{(3)}$ in der Form $\delta_j = \frac{s_j}{t_j}$ dar, wobei jeweils s_j nichtnegativ ganz und t_j positiv ganz ist sowie s_j und t_j zueinander relativ prim sind $(j = 1, 2, \ldots, m)$. Jedes t_j besitzt dabei eine eindeutige Faktorzerlegung $t_j = 2^{\lambda_j} u_j$ mit ganzzahligem Exponenten $\lambda_j \geqq 0$ und einem ungeraden ganzzahligen Faktor u_j. Jetzt ist

$u = \prod_{j=1}^{m} u_j$ eine ungerade ganze Zahl, und man erkennt wie in 1., daß $(\mathfrak{A}^{(3)})^{2\frac{u-1}{2}}$ Matrix eines Automorphismus von $\mathfrak{M}^{(2)}$ ist. Somit bewirkt die zur Matrix $\mathfrak{A}^{(4)} = (\mathfrak{A}^{(3)})^{2\frac{u-1}{2}}\mathfrak{A}^{(3)} = (\mathfrak{A}^{(3)})^{u}$ gehörende Abbildung $A^{(4)}$ einen Isomorphismus von $\mathfrak{M}^{(2)*}$ auf $\mathfrak{M}^{(2)}$, und es lassen sich für die Matrix $\mathfrak{A}^{(4)}$ von $A^{(4)}$ nun leicht die in der Behauptung von 2. angegebenen Eigenschaften verifizieren. In der Tat besitzt $\mathfrak{A}^{(4)}$ die Gestalt von $\mathfrak{A}^{(2)}$ mit Ausnahme der Winkelungleichungen. Für die Größen δ_j von $\mathfrak{A}^{(4)}$ gilt weiter: $\delta_j = \frac{\frac{u}{u_j}s_j}{2^{\lambda_j}} - 2\left[\frac{\frac{u}{u_j}s_j}{2^{\lambda_j+1}}\right]$ $(j = 1, 2, \ldots, m)$. Aus diesen Gleichungen folgt aber im Falle $\lambda_j = 0$: δ_j ist ganzzahlig, d.h. wegen $0 \leqq \delta_j < 2$ entweder $\delta_j = 0$ oder $\delta_j = 1 = \frac{l_j}{2^{\lambda_j}}$ mit $l_j = 1$; während im Falle $\lambda_j > 0$ s_j notwendig ungerade ist und damit die Darstellung $\delta_j = \frac{l_j}{2^{\lambda_j}}$ mit wiederum ungeradem $l_j = \frac{u}{u_j}s_j - 2^{\lambda_j+1}\left[\frac{\frac{u}{u_j}s_j}{2^{\lambda_j+1}}\right]$ gilt. In beiden Fällen besteht für l_j die Ungleichung $0 < l_j < 2^{\lambda_j+1}$.

3. Zur Vollendung des Beweises von Satz 10 benötigen wir noch die folgende Tatsache:

Die Matrix $\mathfrak{A}^{(4)}$ ist mit einer Matrix $\mathfrak{A}^{(5)}$ kongruent, welche die Gestalt von $\mathfrak{A}^{(2)}$ mit unter Umständen mehr einreihigen Hauptminoren (1) und (−1) besitzt und deren Winkel den in Satz 10 angegebenen Bedingungen genügen.

Wir zeigen zuerst, daß $\mathfrak{A}^{(4)}$ mit einer Matrix $\mathfrak{A}_1^{(4)}$ von der Gestalt von $\mathfrak{A}^{(4)}$ kongruent ist, für die außerdem die etwas verschärften Ungleichungen $0 < l_j \leqq 2^{\lambda_j}$ gültig sind. Zu diesem Zweck nehmen wir einmal an, bei einem gewissen Index j_0 gelte für die Größe δ_{j_0} von $\mathfrak{A}^{(4)}$ die Darstellung $\delta_{j_0} = \frac{l_{j_0}}{2^{\lambda_{j_0}}}$ mit der entgegengesetzten Ungleichung $l_{j_0} > 2^{\lambda_{j_0}}$. Dann besitzt die mit $\mathfrak{A}^{(4)}$ kongruente Matrix $\mathfrak{T}^{(j_0)}\mathfrak{A}^{(4)}\mathfrak{T}^{(j_0)*}$ mit $\mathfrak{T}^{(j_0)} = 1 \dotplus \underbrace{\cdots \dotplus}_{l+2j_0-2} 1 \dotplus (-1) \dotplus 1 \underbrace{\dotplus \cdots \dotplus}_{n-l-2j_0+1} 1$ ebenfalls die Gestalt von $\mathfrak{A}^{(4)}$, und für die δ_{j_0} von $\mathfrak{A}^{(4)}$ entsprechende Größe δ'_{j_0} dieser Matrix gilt: $\delta'_{j_0} = \frac{l'_{j_0}}{2^{\lambda_{j_0}}}$, wobei $l'_{j_0} = 2^{\lambda_{j_0}+1} - l_{j_0}$ eine ungerade ganze Zahl mit $0 < l'_{j_0} < 2^{\lambda_{j_0}}$ ist. Indem man nun diese Abänderung entsprechend für alle Indizes j vornimmt, wo sie nötig ist, erhält man die gewünschte, mit $\mathfrak{A}^{(4)}$ kongruente Matrix $\mathfrak{A}_1^{(4)}$. Nun kann man die beiden, für $\mathfrak{A}_1^{(4)}$ noch möglichen Grenzfälle $\delta_j = 0$ und $l_j = 2^{\lambda_j}$ d.h. $\delta_j = 1$ ausschließen, wenn man beachtet, daß die zugehörigen Rotationsmatrizen $\mathfrak{d}_j$ dann die Gestalt $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ und $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ annehmen. Man braucht dazu nur die Minoren $\mathfrak{d}_j$ der Matrix $\mathfrak{A}_1^{(4)}$ so umzuordnen, daß diese speziellen Rotationsmatrizen direkt an die einreihigen Hauptminoren (1) und (−1) von $\mathfrak{A}_1^{(4)}$ anschließen. Die auf diese Weise umgeordnete Matrix $\mathfrak{A}_2^{(4)}$ besitzt dann wieder die Gestalt von $\mathfrak{A}_1^{(4)}$, allerdings mit möglicherweise mehr einreihigen Hauptminoren (1) und (−1), und für ihre Winkel φ_j

gilt stets $\varphi_j = \frac{l_j}{2^{\lambda_j}}\pi$ mit $0<l_j<2^{\lambda_j}$. Durch eine weitere Umordnung der Minoren $\mathfrak{d}_j$ von $\mathfrak{A}_2^{(4)}$ entsteht schließlich die Matrix $\mathfrak{A}^{(5)}$ mit den gewünschten Winkelungleichungen $0<\varphi_1\leqq\varphi_2\leqq\cdots\leqq\varphi_m<\pi$. Aus der Matrizenrelation $\begin{pmatrix} 0 & \mathfrak{d}_0 \\ \mathfrak{d}_0 & 0 \end{pmatrix}\begin{pmatrix} \mathfrak{d}_j & 0 \\ 0 & \mathfrak{d}_{j'} \end{pmatrix}\begin{pmatrix} 0 & \mathfrak{d}_0 \\ \mathfrak{d}_0 & 0 \end{pmatrix} = \begin{pmatrix} \mathfrak{d}_{j'} & 0 \\ 0 & \mathfrak{d}_j \end{pmatrix}$ $\left(\mathfrak{d}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};\ 1\leqq j, j'\leqq m\right)$ entnimmt man aber, daß durch die Umordnungen der $\mathfrak{d}_j$ stets kongruente Matrizen entstehen, so daß also $\mathfrak{A}^{(5)}$ mit $\mathfrak{A}_1^{(4)}$ und somit auch mit $\mathfrak{A}^{(4)}$ kongruent ist, womit die Behauptung von 3. erwiesen ist.

Aus 2. und 3. resultiert jetzt, daß $\mathfrak{M}^{(2)}=T_2\mathfrak{M}$ einen Isomorphismus $A^{(4)}$ mit der Matrix $\mathfrak{A}^{(4)}=\mathfrak{T}_1\mathfrak{A}^{(5)}\mathfrak{T}_1^*$ besitzt ($\mathfrak{T}_1$ ist nichtsingulär). Nach Satz 5 besitzt also $\mathfrak{M}=T_2^{-1}\mathfrak{M}^{(2)}$ den Isomorphismus $A^{(0)}$ mit der Matrix $\mathfrak{A}^{(0)}=$ $=\mathfrak{T}_2^{-1}\mathfrak{A}^{(4)}(\mathfrak{T}_2^{-1})^*=\mathfrak{T}\mathfrak{A}^{(5)}\mathfrak{T}^*$, wobei $\mathfrak{T}=\mathfrak{T}_2^{-1}\mathfrak{T}_1$ eine nichtsinguläre Matrix darstellt. Damit ist aber Satz 10 vollständig bewiesen; w. z. b. w.

Wir weisen an dieser Stelle darauf hin, daß ein selbstadjungierter Minkowski-Raum sehr viele verschiedene kanonische Isomorphismen besitzen kann, wie das Beispiel eines euklidischen Raums zeigt, wo alle nach Satz 10 überhaupt möglichen Matrizen $\mathfrak{A}^{(0)}$ (mit festem $\mathfrak{T}$) Matrizen kanonischer Isomorphismen sind. Man kann sich nun fragen: *Gibt es auch nichttriviale, d. h. von den euklidischen Räumen verschiedene selbstadjungierte Minkowski-Räume mit einem Isomorphismus, dessen Matrix gleich einer beliebig vorgegebenen Matrix der Gestalt von $\mathfrak{A}^{(0)}$ ist?* Zur Beantwortung dieser Frage beweisen wir zunächst:

Satz 11: *Jede Matrix von der Gestalt der in Satz 10 vorkommenden Matrix $\mathfrak{A}^{(0)}$ mit nichtsingulärer Matrix $\mathfrak{T}$ sowie mit l einreihigen Hauptminoren (1), $n-l-2m$ einreihigen Hauptminoren (-1) und m Winkeln $\varphi_j=\frac{l_j}{2^{\lambda_j}}\pi$ $(0<\varphi_j<\pi)$ ist Matrix eines kanonischen Isomorphismus des selbstadjungierten Minkowski-Raums $T(\underbrace{\mathfrak{M}_1^{(1)}\times\cdots\times\mathfrak{M}_1^{(1)}}_{l\text{ Faktoren}}\times\mathfrak{M}_2^{(1)}\times\cdots\times\mathfrak{M}_2^{(m)}\times\underbrace{\mathfrak{M}_1^{(-1)}\times\cdots\times\mathfrak{M}_1^{(-1)}}_{n-l-2m\text{ Faktoren}})$.*

Hierin bedeutet:

a) T die der Matrix $\mathfrak{T}$ entsprechende Abbildung[22]*),*

b) $\mathfrak{M}_1^{(1)}$ den eindimensionalen Minkowski-Raum mit $||x||=||(\xi)||=|\xi|$,

c) $\mathfrak{M}_1^{(-1)}$ einen eindimensionalen Minkowski-Raum mit $x=||(\xi)||=\varrho^{-\operatorname{sign}\xi}|\xi|$ für ein reelles $\varrho\geqq 1$, sowie schließlich

d) $\mathfrak{M}_2^{(j)}$ einen zweidimensionalen Minkowski-Raum, dessen durch $||x||=1$ gegebene Punktmenge (euklidisch gesprochen) ein reguläres Polygon mit dem Schwerpunkt $x=0$ ist, welches im Falle $\lambda_j=1$ sechs Ecken und den Durchmesser $2\left(\cos\frac{\pi}{6}\right)^{-1}$ und im Falle $\lambda_j>1$ 2^{λ_j} Ecken und den Durchmesser $2\left(\cos\frac{\pi}{2^{\lambda_j}}\right)^{-1}$ besitzt $(j=1, 2, \ldots, m)$.

Bevor wir diesen Satz beweisen, zeigen wir zunächst:

[22]) Jede derartige Abbildung T bewirkt einen Isomorphismus eines gegebenen Minkowski-Raums auf einen geeignet definierten Minkowski-Raum, wie im ersten Teil des Beweises von Satz 10 gezeigt wurde.

Lemma 6: *Stellt man die Punkte eines beliebigen n-dimensionalen Minkowski-Raums $\mathfrak{M}$ und diejenigen seines Dualraums $\mathfrak{M}^*$ in der auf S. 344 angegebenen Weise in den Koordinaten $\xi_1, \ldots, \xi_n$ eines affinen Raums dar, so sind die Hyperflächen $||x|| = 1$ $(x \in \mathfrak{M})$ und $||f|| = 1$ $(f \in \mathfrak{M}^*)$ polarreziprok bezüglich der durch $\sum_{i=1}^{n} (\xi_i)^2 = 1$ gegebenen Einheitssphäre $\mathfrak{e}$.*

Beweis. Sei f irgendein Punkt der Hyperfläche $||f|| = 1$. Dann ist wegen $||f|| = \sup_{||x||=1} f(x)$ die durch $f(x) = \sum_{i=1}^{n} \alpha_i \xi_i = 1$ definierte Hyperebene eine Stützhyperebene der Hyperfläche $||x|| = 1$. Hierbei sind $\alpha_1, \ldots, \alpha_n$ die Koordinaten des Punktes f, welcher genau den Pol der Hyperebene $f(x) = 1$ bezüglich $\mathfrak{e}$ darstellt. $||f|| = 1$ ist also die Menge der Pole aller Stützhyperebenen von $||x|| = 1$, da man diese Überlegung auch ebenso umkehren kann, oder anders gesagt: $||f|| = 1$ ist die Polarhyperfläche der Hyperfläche $||x|| = 1$ in bezug auf $\mathfrak{e}$; w. z. b. w.

Beweis von Satz 11. Nach Satz 5 und Satz 8 genügt es zu zeigen, daß $\mathfrak{M}_1^{(1)}$, $\mathfrak{M}_1^{(-1)}$ und $\mathfrak{M}_2^{(j)}$ selbstadjungierte Minkowski-Räume sind, welche Isomorphismen mit den Matrizen (1), (−1) und $\mathfrak{d}_j = \begin{pmatrix} \cos\varphi_j & \sin\varphi_j \\ -\sin\varphi_j & \cos\varphi_j \end{pmatrix} \left(\varphi_j = \frac{l_j}{2^{\lambda_j}}\pi\right)$ besitzen. Nun sind aber $\mathfrak{M}_1^{(1)*}$ bzw. $\mathfrak{M}_1^{(-1)*}$ nach Lemma 6 Minkowski-Räume mit $||(\alpha)|| = |\alpha|$ bzw. $||(\alpha)|| = \varrho^{\operatorname{sign}\alpha}|\alpha|$; und nach demselben Lemma findet man, daß die Kurve $||f|| = 1$ $(f \in \mathfrak{M}_2^{(j)*})$ ein Polygon ist, welches aus dem Polygon $||x|| = 1$ $(x \in \mathfrak{M}_2^{(j)})$ durch eine Drehung des letzteren um den Nullpunkt mit dem Winkel $-\frac{l_j}{2^{\lambda_j}}\pi$ hervorgeht $(1 \leqq j \leqq m)$. Daraus folgt in der Tat, daß $\mathfrak{M}_1^{(1)}$, $\mathfrak{M}_1^{(-1)}$ und $\mathfrak{M}_2^{(j)}$ selbstadjungierte Räume sind, welche Isomorphismen mit den angegebenen Matrizen besitzen; w. z. b. w.

Bemerkungen. 1. Natürlich existieren auch selbstadjungierte Minkowski-Räume mit einem kanonischen Isomorphismus $A^{(0)}$, die von dem nach Satz 11 $A^{(0)}$ entsprechenden Produktraum sehr wohl verschieden sind. So hatten wir schon auf S. 344 den Produktraum $\mathfrak{M} \times \mathfrak{M}^*$ als symmetrisch selbstadjungiert kennengelernt, wobei $\mathfrak{M}$ ein beliebiger n'-dimensionaler Minkowski-Raum ist. $\mathfrak{M} \times \mathfrak{M}^*$ besitzt den zur Matrix $\mathfrak{A}^{(0)} = \mathfrak{T}_0 \big(\underbrace{1 \dotplus \cdots \dotplus 1}_{n'} \dotplus (-1) \underbrace{\dotplus \cdots \dotplus (-1)}_{n'}\big) \mathfrak{T}_0^*$ gehörenden kanonischen Isomorphismus ($\mathfrak{T}_0$ bedeutet hierin eine ganz bestimmte nichtsinguläre Matrix) und ist sicher nicht mit dem streng konvexen[23]) Minkowski-Raum $\underbrace{\mathfrak{M}_1^{(1)} \times \cdots \times \mathfrak{M}_1^{(1)}}_{n'} \times \underbrace{\mathfrak{M}_1^{(-1)} \times \cdots \times \mathfrak{M}_1^{(-1)}}_{n'}$ isomorph, wenn $\mathfrak{M}$ geeignet, nämlich etwa als ein nicht streng konvexer Raum gewählt wird. Ebenso ist z. B. der Minkowski-Raum $\mathfrak{S}$, dessen Punktmenge $||x|| \leqq 1$ irgendein n-Simplex mit dem Schwerpunkt $x = 0$ darstellt, symmetrisch selbstadjungiert; er besitzt einen kanonischen Isomorphismus mit der Matrix $\mathfrak{A}^{(0)} = \mathfrak{T}(-\mathfrak{E})\mathfrak{T}^*$, wobei $\mathfrak{T}$ eine nichtsinguläre Matrix und $\mathfrak{E}$ die Einheitsmatrix ist.

[23]) Wir definieren in der üblichen Weise einen linearen schiefnormierten Raum als streng konvex, wenn er keine linear unabhängigen Punkte x und y mit $||x+y|| = ||x|| + ||y||$ enthält.

Als nicht streng konvexer Raum ist $\mathfrak{S}$ wiederum nicht mit dem entsprechenden, streng konvexen Produktraum $\mathfrak{M}_1^{(-1)} \times \cdots \times \mathfrak{M}_1^{(-1)}$ isomorph ($n > 1$).

2. Ein symmetrischer selbstadjungierter Minkowski-Raum besitzt immer einen kanonischen Isomorphismus mit einer symmetrischen Matrix $\mathfrak{A}^{(0)}$ (für die also $m = 0$ gilt), da bekanntlich jede symmetrische, nichtsinguläre Matrix schon von der Gestalt einer derartigen Matrix $\mathfrak{A}^{(0)}$ ist. Wir bemerken, daß die in Satz 11 auftretenden Minkowski-Räume $\mathfrak{M}_1^{(1)}$, $\mathfrak{M}_1^{(-1)}$ und $\mathfrak{M}_2^{(j)}$ alle symmetrisch sind, was damit nach Satz 8 auch für ihre Produkträume zutrifft. Um dies für $\mathfrak{M}_2^{(j)}$ einzusehen, braucht man nämlich nur Lemma 6 anzuwenden und zu beachten, daß das Polygon $\|f\| = 1$ von $\mathfrak{M}_2^{(j)*}$ auch durch eine geeignete Spiegelung an einer durch den Nullpunkt gehenden Geraden in das Polygon $\|x\| = 1$ von $\mathfrak{M}_2^{(j)}$ übergeführt werden kann und daß die Matrix jeder derartigen Spiegelung symmetrisch ist ($j = 1, 2, \ldots, m$). Übrigens existieren auch selbstadjungierte Minkowski-Räume, die nicht symmetrisch sind[24]).

Der Produktraum von Satz 11 ist dann und nur dann euklidisch, wenn $m = 0$ und $\varrho = 1$ gilt. Er ist außerdem dann und nur dann von vollhomogener Norm, wenn $\varrho = 1$ ist. Daraus folgt aber in Verbindung mit Satz 11 als Antwort auf die auf S. 350 gestellte Frage, *daß sich jede Matrix von der Gestalt von $\mathfrak{A}^{(0)}$, für die $l < n$ ist, als Matrix eines kanonischen Isomorphismus eines nichttrivialen selbstadjungierten Minkowski-Raums realisieren läßt. Ebenso läßt sich auch jede Matrix von der Gestalt von $\mathfrak{A}^{(0)}$, bei welcher $l < n$ und $n - l - 2m < n$ gilt, in einem nichttrivialen selbstadjungierten Minkowski-Raum mit vollhomogener Norm realisieren.* Im Falle $m > 0$ ist dies wieder eine Folge von Satz 11, und im Falle $m = 0$ kann man auf Grund von Bemerkung 2)

$$T\,(\underbrace{\mathfrak{M}_1^{(1)} \times \cdots \times \mathfrak{M}_1^{(1)}}_{l-1} \times \mathfrak{M}_2 \times \underbrace{\mathfrak{M}_1^{(1)} \times \cdots \times \mathfrak{M}_1^{(1)}}_{n-l-1})$$

als geeigneten selbstadjungierten Minkowski-Raum wählen, worin $\mathfrak{M}_2$ irgendeiner der in Satz 11 definierten Räume $\mathfrak{M}_2^{(j)}$ ist. Wir werden im nächsten Paragraphen sehen, daß im Gegensatz dazu eine Matrix der Gestalt $\mathfrak{A}^{(0)}$ mit $l = n$ bzw. mit $l = n$ oder $n - l - 2m = n$ in einem nichttrivialen selbstadjungierten Minkowski-Raum bzw. in einem nichttrivialen selbstadjungierten Minkowski-Raum mit vollhomogener Norm nicht realisiert werden kann.

§ 5. Zur Kennzeichnung der euklidischen Räume unter den selbstadjungierten Minkowski-Räumen

Wir notieren zunächst:

DEFINITION 15. *Eine n-reihige, symmetrische Matrix $\mathfrak{A}$ mit den Elementen α_{ik} wird positiv bzw. negativ definit genannt, wenn die zugehörige quadratische Form $\sum_{i=1}^{n} \sum_{k=1}^{n} \alpha_{ik} \xi_i \xi_k$ positiv bzw. negativ definit ist.*

[24]) In einer Fortsetzung zu dieser Arbeit werden dafür Beispiele angegeben werden.

Dann läßt sich eine erste, einigermaßen triviale Kennzeichnung der euklidischen Räume in der folgenden Weise ausdrücken:

Satz 12. *Ein selbstadjungierter Minkowski-Raum $\mathfrak{M}$ ist dann und nur dann ein euklidischer Raum, wenn er einen Isomorphismus A mit einer positiv definiten symmetrischen Matrix $\mathfrak{A}$ besitzt.*

Beweis. Die Notwendigkeit der angegebenen Bedingung ist wegen Definition 11 und Satz 5 evident, wonach ein euklidischer Raum einen Isomorphismus mit der positiv definiten symmetrischen Matrix $\mathfrak{A}=\mathfrak{T}\mathfrak{E}\mathfrak{T}^*$ besitzt hierbei ist $\mathfrak{T}$ nichtsingulär und $\mathfrak{E}$ die Einheitsmatrix). Es werde nun umgekehrt angenommen, der selbstadjungierte Minkowski-Raum $\mathfrak{M}$ besitze ein A mit positiv definiter symmetrischer Matrix $\mathfrak{A}$. Dann gilt $\mathfrak{A}=\mathfrak{T}\mathfrak{E}\mathfrak{T}^*$ mit geeignetem nichtsingulären $\mathfrak{T}$, dessen zugehörige Abbildung mit T bezeichnet sei. Nach Satz 5 ist $\mathfrak{M}^{(1)}=T^{-1}\mathfrak{M}$ ein selbstadjungierter Raum mit einem Isomorphismus, dessen Matrix die Einheitsmatrix $\mathfrak{E}$ ist (vgl. Fußnote [22]). Daraus folgt aber die Identität der Hyperflächen $||x||=1$ ($x\in\mathfrak{M}^{(1)}$) und $||f||=1$ ($f\in\mathfrak{M}^{(1)*}$) in dem n-dimensionalen affinen Raum, in welchem sich diese gemeinsam darstellen lassen. Wir betrachten jetzt einen Punkt x_0 von $||x||=1$, dessen gewöhnlicher euklidischer Abstand vom Nullpunkt im Vergleich zu dem Abstand der übrigen Punkte dieser Hyperfläche vom Nullpunkt maximal ist. Dann geht durch x_0 eine Stützhyperebene h_0 an $||x||=1$, welche zu der von dem Nullpunkt ausgehenden und x_0 passierenden Halbgeraden d' (im gewöhnlichen Sinne) orthogonal ist. Die Gleichung dieser Hyperebene kann in der Form $f_0(x)=1$ geschrieben werden, wobei f_0 gerade ihren Pol bezüglich der Einheitssphäre $\mathfrak{e}$ unseres affinen Raums darstellt. f_0 liegt nach Lemma 6 auf der Hyperfläche $||f||=1$ und damit auch auf der mit ihr zusammenfallenden Hyperfläche $||x||=1$. Andererseits ist wegen der Orthogonalität von h_0 und d' f_0 als Pol von h_0 in d' enthalten, d.h. f_0 ist mit dem Schnittpunkt x_0 von $||x||=1$ und d' identisch. Also ist x_0 ein Punkt seiner eigenen Polarhyperebene h_0 und somit in $\mathfrak{e}$ enthalten. Auf dieselbe Weise findet man, daß auch jeder Punkt von $||x||=1$ mit minimalem euklidischen Abstand vom Nullpunkt auf $\mathfrak{e}$ liegt. Daraus folgt aber die Identität von $||x||=1$ und $\mathfrak{e}$ und damit die Tatsache, daß der Raum $\mathfrak{M}^{(1)}$ sowie auch der Raum $\mathfrak{M}$ euklidisch ist; w.z.b.w.

Wir erwähnen, daß sich Satz 12 mit Hilfe des auf S. 339 eingeführten inneren Produktes auch in der folgenden Weise ausdrücken läßt: *Ein Minkowski-Raum ist dann und nur dann euklidisch, wenn in ihm eine Bilinearform $\langle x, y\rangle$ mit $\sup_{||y||=1}\langle x, y\rangle=||x||$ definiert werden kann, welche symmetrisch ist und für die die quadratische Form $\langle x, x\rangle$ positiv definit ist*[25]). Da ein Minkowski-Raum mit vollhomogener Norm immer die Spiegelung am Nullpunkt mit der Matrix $-\mathfrak{E}$ als Automorphismus gestattet, folgt aus Satz 12 als

Korollar. *Ein selbstadjungierter Minkowski-Raum mit vollhomogener Norm ist dann und nur dann euklidisch, wenn er einen Isomorphismus A mit*

[25]) Man kann übrigens zeigen, daß aus dem Erfülltsein dieser Bedingungen für das innere Produkt $\langle x, y\rangle$ automatisch $\langle x, x\rangle=||x||^2$ folgt.

einer positiv definiten oder einer negativ definiten, symmetrischen Matrix $\mathfrak{A}$ *besitzt.*

BEMERKUNG. Eine direkte Folge von Satz 12 und dem Korollar ist die schon am Ende von § 4 erwähnte Tatsache, daß eine Matrix der Gestalt von dem in Satz 10 vorkommenden $\mathfrak{A}^{(0)}$ mit $l=n$ nicht Matrix eines kanonischen Isomorphismus eines nichttrivialen selbstadjungierten Minkowski-Raums sein kann und daß sich ebenso eine Matrix der Gestalt von $\mathfrak{A}^{(0)}$ mit $l=n$ oder $n-l-2m=n$ in einem nichttrivialen selbstadjungierten Minkowski-Raum mit vollhomogener Norm nicht realisieren läßt.

Die Bedingung von Satz 12 kann nicht in der Weise abgeschwächt werden, daß schon die Existenz eines Isomorphismus A mit einer solchen Matrix $\mathfrak{A}$ für einen euklidischen Raum charakteristisch wäre, deren symmetrisierte Matrix $\frac{1}{2}(\mathfrak{A}+\mathfrak{A}^*)$ positiv definit ist. Dies zeigt nämlich das Beispiel der in Satz 11 auftretenden Räume $\mathfrak{M}_2^{(j)}$ im Fall $\lambda_j>1$, die jeweils ein A mit $\mathfrak{A}=\begin{pmatrix}\cos\varphi_j & \sin\varphi_j\\ -\sin\varphi_j & \cos\varphi_j\end{pmatrix}$ und $\varphi_j=\frac{\pi}{2^{\lambda_j}}<\frac{\pi}{2}$ besitzen. Nimmt man aber noch eine geeignete Bedingung für die übrigen, auf S. 347 angegebenen invarianten Größen des Isomorphismus A hinzu, so kann man zeigen:

Satz 13: *Ein selbstadjungierter Minkowski-Raum* $\mathfrak{M}$ *ist dann und nur dann ein euklidischer Raum, wenn er (mindestens) einen Isomorphismus* A *besitzt, dessen Matrix* $\mathfrak{A}$ *die folgenden Eigenschaften hat:*

a) $\frac{1}{2}(\mathfrak{A}+\mathfrak{A}^*)$ *ist positiv definit, und*

b) die Kongruenzinvarianten $\delta_j=\frac{\varphi_j}{\pi}$ *von* $\mathfrak{A}$[26]) $(j=1, 2, \ldots, m)$ *sind zusammen mit der Zahl 1 über dem Körper der rationalen Zahlen linear unabhängig.*

BEWEIS. Die Notwendigkeit der angegebenen Bedingungen ist offensichtlich, da ein euklidischer Raum stets einen Isomorphismus A besitzt, dessen symmetrische Matrix $\mathfrak{A}=\mathfrak{T}\mathfrak{E}\mathfrak{T}^*$ selbst positiv definit ist und für welche $m=0$ gilt, so daß hierbei Bedingung b) trivialerweise erfüllt ist. Um ihr Hinreichen einzusehen, bemerken wir, daß der Minkowski-Raum $\mathfrak{M}$ auf Grund von Satz 9, Lemma 3, Satz 5 und den in Satz 13 formulierten Bedingungen mit einem selbstadjungierten Minkowski-Raum $\mathfrak{M}^{(2)}$ isomorph ist, der einen Isomorphismus mit der Matrix $\mathfrak{A}^{(2)}=\underbrace{1\dotplus\cdots\dotplus 1}_{l}\dotplus\mathfrak{d}_1\dotplus\cdots\dotplus\mathfrak{d}_m$ besitzt, wobei $\mathfrak{d}_j=\begin{pmatrix}\cos\varphi_j & \sin\varphi_j\\ -\sin\varphi_j & \cos\varphi_j\end{pmatrix}$ und $0<\varphi_j<\frac{\pi}{2}$ gilt $(j=1, 2, \ldots m)$. Wegen Satz 6 sind nun $(\mathfrak{A}^{(2)})^2$ und damit auch alle Potenzen $(\mathfrak{A}^{(2)})^{2\nu}$ mit beliebigem ganzzahligem ν Matrizen von Automorphismen des Raums $\mathfrak{M}^{(2)}$. Daraus folgt aber durch Anwendung von Lemma 5 für alle δ_j und 1 analog wie in Teil 1. des Beweises von Satz 10 die Existenz eines Automorphismus von $\mathfrak{M}^{(2)}$ mit der Matrix $\mathfrak{A}^{(2)}$. Also ist $\mathfrak{E}=(\mathfrak{A}^{(2)})^{-1}\mathfrak{A}^{(2)}$ die positiv definite symmetrische Matrix eines Isomorphismus von $\mathfrak{M}^{(2)*}$ auf $\mathfrak{M}^{(2)}$, d.h. nach Satz 12 ist $\mathfrak{M}^{(2)}$ euklidisch, was damit auch für den mit $\mathfrak{M}^{(2)}$ isomorphen Raum $\mathfrak{M}$ der Fall ist; w.z.b.w.

[26]) Diese Größen lassen sich nach Satz 9 und Lemma 3 in (wegen Lemma 4 und Bedingung a)) eindeutiger Weise aus $\mathfrak{A}$ gewinnen.

Es läßt sich übrigens unter Benutzung einer Verschärfung von Lemma 5[27]) zeigen, daß auch die folgende Bedingung ein hinreichendes Kriterium dafür darstellt, daß der selbstadjungierte Raum $\mathfrak{M}$ euklidisch ist: $\mathfrak{M}$ besitzt (mindestens) einen Isomorphismus A, dessen Matrix $\mathfrak{A}$ die Eigenschaften hat, daß $\frac{1}{2}(\mathfrak{A}+\mathfrak{A}^*)$ positiv definit ist und daß die von $\delta_1, \delta_2, \ldots, \delta_m, 1$ bzw. von $\delta_1, \delta_2, \ldots, \delta_m$ im Körper der reellen Zahlen erzeugten Vektorräume mit dem Körper R aller rationalen Zahlen als Skalarbereich $R(\delta_1, \ldots, \delta_m, 1)$ bzw. $R(\delta_1, \ldots, \delta_m)$ die Dimensionen m bzw. $m-1$ besitzen, wobei δ_j die in Satz 13 angegebenen Invarianten von $\mathfrak{A}$ sind ($j=1, 2, \ldots, m$).

Während Satz 13 eine arithmetische Charakterisierung eines euklidischen Raums ausdrückt, wird durch den nachfolgenden Satz eine geometrische Kennzeichnung der euklidischen Räume mit einer Dimension größer als zwei gegeben:

Satz 14: *Ein selbstadjungierter Minkowski-Raum $\mathfrak{M}$ mit vollhomogener Norm und einer Dimension größer als zwei ist dann und nur dann ein euklidischer Raum, wenn er mit einem Minkowski-Raum $\mathfrak{M}^{(1)}$ isomorph ist, welcher alle (im gewöhnlichen Sinne) eigentlichen euklidischen Bewegungen als Automorphismen gestattet, die eine durch den Nullpunkt gehende Gerade d_1 punktweise festlassen*[28]).

Beweis. Wenn $\mathfrak{M}$ euklidisch ist, so ist seine Isomorphie mit einem im gewöhnlichen Sinne rotationssymmetrischen Minkowski-Raum trivial. Wir nehmen jetzt an, $\mathfrak{M}$ sei selbstadjungiert, von vollhomogener Norm und von der Dimension $n>2$. Außerdem gelte $\mathfrak{M} \cong \mathfrak{M}^{(1)}$ für einen Minkowski-Raum $\mathfrak{M}^{(1)}$ mit der in unserer Behauptung geforderten Eigenschaft. Wir bezeichnen die Gruppe der Matrizen aller Automorphismen von $\mathfrak{M}$ bzw. $\mathfrak{M}^{(1)}$ mit G bzw. G_1. Nun folgt aus einem Satz von J. v. Neumann wegen der gleichmäßigen Beschränktheit aller Elemente der Matrizen aus G die Existenz einer solchen nichtsingulären Matrix $\mathfrak{T}$, daß $G_2=\mathfrak{T}G\mathfrak{T}^{-1}$ eine Untergruppe der Gruppe O_n aller n-reihigen, orthogonalen Matrizen darstellt (s. Fußnote [17])). Offensichtlich ist G_2 die Gruppe der Matrizen aller Automorphismen des Raumes $\mathfrak{M}^{(2)}=T\mathfrak{M}$, wenn T die $\mathfrak{T}$ entsprechende Abbildung ist. $\mathfrak{M}^{(2)}$ ist mit $\mathfrak{M}$ und somit auch mit $\mathfrak{M}^{(1)}$ isomorph, d. h. es gibt genauer gesagt eine Abbildung T_1 mit der nichtsingulären Matrix $\mathfrak{T}_1$, so daß $\mathfrak{M}^{(2)}=T_1\mathfrak{M}^{(1)}$ und $G_2=\mathfrak{T}_1 G_1 \mathfrak{T}_1^{-1}$ gilt. Es sei weiter $d=T_1 d_1$ und $(SO)_n^{(d)}$ bzw. $(SO)_n^{(d_1)}$ die Gruppe der Matrizen aller derjenigen eigentlichen euklidischen Bewegungen des n-dimensionalen Raums, die die Gerade d bzw. d_1 punktweise festlassen. Dann besagt unsere Voraussetzung über $\mathfrak{M}^{(1)}$: $(SO)_n^{(d_1)} \subseteqq G_1$; und daraus folgt wegen $\mathfrak{T}_1 (SO)_n^{(d_1)} \mathfrak{T}_1^{-1} \subseteqq G_2 \subseteqq O_n$ und $d=T_1 d_1$: $\mathfrak{T}_1 (SO)_n^{(d_1)} \mathfrak{T}_1^{-1} \subseteqq (SO)_n^{(d)}$. Die Anwendung des Brouwerschen Satzes von der Gebietsinvarianz einer topologischen Abbildung[29]) auf die mit der trivialen Topologie von O_n versehene Mannigfaltigkeit $(SO)_n^{(d_1)}$ und die (bezüglich derselben Topologie) zusammenhängende Mannigfaltigkeit $(SO)_n^{(d)}$[30])

[27]) Es handelt sich hierbei um Satz 4 von [13], S. 83.

[28]) Das heißt anders ausgedrückt, wenn $\mathfrak{M}$ mit einem bezüglich d_1 rotationssymmetrischen $\mathfrak{M}^{(1)}$ isomorph ist.

[29]) Siehe [14], S. 396, Satz 9.

[30]) Hierbei wird die Voraussetzung $n>2$ benutzt.

zeigt nun die Gleichheit von $\mathfrak{T}_1(SO)_n^{(d_1)}\mathfrak{T}_1^{-1}$ und $(SO)_n^{(d)}$, woraus $G_2 \supseteq (SO)_n^{(d)}$ resultiert.

Wir betrachten jetzt neben G_2 noch die Gruppe der Matrizen aller Automorphismen von $\mathfrak{M}^{(2)*}$, welche mit G_2^* bezeichnet sei. G_2^* enthält ebenfalls $(SO)_n^{(d)}$, weil die zu einem beliebigen Element von $(SO)_n^{(d)}$ gehörende Abbildung als Automorphismus von $\mathfrak{M}^{(2)}$ nicht nur die Gesamtheit aller Stützhyperebenen von $||x|| = 1$ $(x \in \mathfrak{M}^{(2)})$ in sich überführt, sondern als Bewegung auch die, nach Lemma 6 gerade die Punkte von $||f|| = 1$ $(f \in \mathfrak{M}^{(2)*})$ ausmachenden Pole dieser Hyperebenen bezüglich der Einheitssphäre $\mathfrak{e}$ (als Gesamtheit) in sich abbildet. Weiter gilt $G_2 = \mathfrak{A} G_2^* \mathfrak{A}^{-1}$, wenn $\mathfrak{A}$ die Matrix eines Isomorphismus A von $\mathfrak{M}^{(2)*}$ auf $\mathfrak{M}^{(2)}$ ist; und hieraus folgt wegen $\mathfrak{A}(SO)_n^{(d)}\mathfrak{A}^{-1} \subseteq G_2 \subseteq O_n$: $\mathfrak{A}(SO)_n^{(d)}\mathfrak{A}^{-1} \subseteq (SO)_n^{(Ad)}$, wobei $(SO)_n^{(Ad)}$ analog wie $(SO)_n^{(d)}$ erklärt ist. Schließlich können wir wiederum unter Benutzung des Satzes der Gebietsinvarianz auf die Gleichheit in der letzten Inklusion schließen und erhalten somit $G_2 \supseteq (SO)_n^{(Ad)}$.

Jetzt müssen zwei Fälle unterschieden und gesondert diskutiert werden, nämlich 1. $Ad \neq d$ und 2. $Ad = d$. Setzen wir also zuerst einmal voraus, daß die (durch den Nullpunkt gehende) Gerade Ad von der (ebenfalls durch den Nullpunkt gehenden) Geraden d verschieden sei. Dann erzeugen wegen $n \geqq 3$ die beiden Untergruppen $(SO)_n^{(d)}$ und $(SO)_n^{(Ad)}$ von O_n schon die Gruppe $(SO)_n$ aller n-reihigen, orthogonalen Matrizen der Determinante 1, was man durch Zurückführung auf den analogen Sachverhalt in einem dreidimensionalen, d und Ad enthaltenden Unterraum einsehen kann, wenn man beachtet, daß es hierbei nur auf den Nachweis der Transitivität der zu der von $(SO)_n^{d}$ und $(SO)_n^{Ad}$ erzeugten Gruppe gehörenden Abbildungen auf $\mathfrak{e}$ ankommt. Da aber G_2 nach dem bisher Bewiesenen $(SO)_n^{d}$ und $(SO)_n^{Ad}$ enthält, folgt $G_2 \supseteq (SO)_n$ und damit die Tatsache, daß $\mathfrak{M}^{(2)}$ wie auch $\mathfrak{M}$ euklidisch sein müssen, womit unser Beweis in diesem Falle beendet ist.

Im Falle $Ad = d$ können wir stets zusätzlich erreichen, daß A auch eine vom Nullpunkt ausgehende Halbgerade d' von d als Ganzes invariant läßt, indem wir unter Benutzung der Voraussetzung, daß $\mathfrak{M}$ und deshalb auch $\mathfrak{M}^{(2)}$ Minkowski-Räume mit vollhomogener Norm sind, notfalls A durch das Produkt von A und der Spiegelung am Nullpunkt ersetzen. Wir behaupten jetzt, daß die Hyperfläche $||x|| = 1$ von $\mathfrak{M}^{(2)}$ in ihrem Schnittpunkt x_0 mit d' eine Tangentialhyperebene h_0, d.h. eine eindeutig bestimmte Stützhyperebene besitzt, welche zu d' (im gewöhnlichen Sinn) orthogonal ist. Wäre nämlich $h_0^{(1)}$ eine durch x_0 gehende und nicht zu d' orthogonale Stützhyperebene von $||x|| = 1$, so enthielte $h_0^{(1)}$ sicher eine von x_0 ausgehende Halbgerade, die mit der Strecke $\overline{x_0 0}$ einen Winkel kleiner als $\pi/2$ bildet und auf der kein Punkt x aus $\mathfrak{M}^{(2)}$ mit $||x|| < 1$ liegt. Dann hätte aber der durch alle Rotationen dieser Halbgeraden um d erzeugte Hyperkegel $\mathfrak{k}$ wegen der Rotationssymmetrie der Hyperfläche $||x|| = 1$ bezüglich d die Eigenschaft, daß jede durch seine Spitze x_0 gehende Stützhyperebene von $\mathfrak{k}$ auch Stützhyperebene von $||x|| = 1$ ist. Auf Grund der in Lemma 6 festgestellten Polarreziprozität von $||x|| = 1$ $(x \in \mathfrak{M}^{(2)})$

und $||f|| = 1$ ($f \in \mathfrak{M}^{(2)*}$) bedeutete dies, daß die Hyperfläche $||f|| = 1$ von $\mathfrak{M}^{(2)*}$ in einer Umgebung ihres Schnittpunkts f_0 mit d' einem Hyperebenenstück gleicht. Der Isomorphismus A, welcher wegen $A\,d' = d'$ f_0 in x_0 überführt, würde jetzt dieses Hyperebenenstück auf ein in einer Umgebung von x_0 gelegenes Hyperebenenstück von $||x|| = 1$ abbilden, was mit der Existenz von f im Widerspruch steht. Damit ist aber gezeigt, daß die zu d' senkrechte und durch x_0 gehende Hyperebene h_0 Tangentialhyperebene von $||x|| = 1$ ist. Auf dieselbe Weise sieht man ein, daß auch die zu d' senkrechte und durch f_0 gehende Hyperebene k_0 Tangentialhyperebene von $||f|| = 1$ ist. Die zentralaffine Abbildung A führt also k_0 in h_0 über; und deshalb gilt für die mit k_0 und h_0 parallele, den Nullpunkt enthaltende Hyperebene h: $A\,h = h$. A induziert in dieser zu d orthogonalen Hyperebene h eine nichtausgeartete Affinität, die den Durchschnitt von h mit $||f|| = 1$, eine $(n-2)$-dimensionale Sphäre, in den ebenfalls sphärenförmigen Durchschnitt von h mit $||x|| = 1$ überführt und somit eine Ähnlichkeitsabbildung darstellt.

Die bis jetzt bewiesenen Eigenschaften von A lassen die folgende analytische Formulierung zu: Die Matrix $\mathfrak{A}$ von A besitzt die Darstellung $\mathfrak{A} = \mathfrak{O}\begin{pmatrix} \sigma & 0 \\ 0 & \tau\mathfrak{O}_{n-1} \end{pmatrix}\mathfrak{O}^{-1}$, wobei σ und τ jeweils eine positive reelle Zahl, $\mathfrak{O}_{n-1}$ eine $(n-1)$-reihige orthogonale Matrix und $\mathfrak{O}$ eine n-reihige orthogonale Matrix bedeutet, deren zugehörige Bewegung die ξ_1-Achse in d abbildet. Nun ist aber $\mathfrak{B}_2 = \mathfrak{O}\begin{pmatrix} 1 & 0 \\ 0 & \mathfrak{O}_{n-1} \end{pmatrix}\mathfrak{O}^{-1}$ die Matrix einer Bewegung, die d punktweise festläßt und aus diesem Grunde wegen der Rotationssymmetrie von $\mathfrak{M}^{(2)}$ einen Automorphismus dieses Raums bewirkt. Also stellt $\mathfrak{B}_2^{-1}\mathfrak{A} = \mathfrak{O}\left(\sigma \dotplus \underbrace{\tau \dotplus \cdots \dotplus \tau}_{n-1}\right)\mathfrak{O}^*$ die positiv definite symmetrische Matrix eines Isomorphismus von $\mathfrak{M}^{(2)*}$ auf $\mathfrak{M}^{(2)}$ dar. Nach Satz 12 ist deshalb $\mathfrak{M}^{(2)}$ sowie der damit isomorphe Raum $\mathfrak{M}$ tatsächlich euklidisch, und damit ist alles bewiesen; w.z.b.w.

Satz 14 läßt sich ohne die Voraussetzung einer vollhomogenen Norm in der folgenden Weise formulieren:

KOROLLAR. *Ein selbstadjungierter Minkowski-Raum mit einer Dimension größer als zwei ist dann und nur dann euklidisch, wenn er mit einem Minkowski-Raum $\mathfrak{M}^{(1)}$ isomorph ist, welcher alle eigentlichen euklidischen Bewegungen als Automorphismen gestattet, die eine durch den Nullpunkt gehende Gerade d_1 als Ganzes und den Nullpunkt selbst festlassen.*

Um dies einzusehen, genügt es zu zeigen, daß ein Minkowski-Raum mit der in diesem Korollar angegebenen Rotationssymmetrie notwendig eine vollhomogene Norm besitzt. In der Tat ist die eigentliche Bewegung, die in der durch d_1 und irgendeinen Punkt $x \in \mathfrak{M}^{(1)}$ gehenden Ebene eine Halbdrehung um den Nullpunkt und in dem orthogonalen Komplement davon die identische Abbildung induziert, ein Automorphismus von $\mathfrak{M}^{(1)}$, welcher x in $-x$ überführt, so daß $||x|| = ||-x||$ folgt.

Für die Dimension $n = 1$ ist die Behauptung von Satz 14 ebenfalls trivialerweise richtig, aber für $n = 2$ wird sie falsch, da dann die identische Abbildung die einzige eigentliche Bewegung ist, die d_1 punktweise festläßt, und

sehr wohl in den Räumen $\mathfrak{W}_2^{(j)}$ von Satz 11 nichttriviale selbstadjungierte Minkowski-Ebenen mit vollhomogener Norm existieren. Jetzt kann man aber Satz 14 den folgenden Satz an die Seite stellen:

Satz 15: *Ein zwei- oder dreidimensionaler, selbstadjungierter Minkowski-Raum $\mathfrak{M}$ mit vollhomogener Norm ist dann und nur dann euklidisch, wenn er unendlich viele verschiedene Isomorphismen A besitzt.*

BEWEIS. Wie wir schon am Anfang des Beweises von Satz 14 erläuterten, ist $\mathfrak{M}$ mit einem Raum $\mathfrak{M}^{(1)}$ isomorph, dessen sämtliche Automorphismen Bewegungen sind. Die Gruppe Γ_1' aller Automorphismen von $\mathfrak{M}^{(1)}$ ist also eine in bezug auf die natürliche Topologie kompakte und — nach der Voraussetzung von unendlich vielen verschiedenen Isomorphismen von $\mathfrak{M}^*$ auf $\mathfrak{M}$, d.h. wegen Satz 5 von ebenfalls unendlich vielen verschiedenen Isomorphismen von $\mathfrak{M}^{(1)*}$ auf $\mathfrak{M}^{(1)}$ — unendliche Untergruppe der Gruppe aller (den Nullpunkt festlassenden) Bewegungen des zwei- bzw. dreidimensionalen Raums. Man kann noch weitergehend feststellen, daß analog dazu die Gruppe Γ_1^+ aller der Automorphismen von $\mathfrak{M}^{(1)}$, deren zugehörige Matrix eine positive Determinante besitzt, eine kompakte und unendliche Untergruppe der Gruppe aller Rotationen der Ebene bzw. des dreidimensionalen Raums um den Nullpunkt darstellt. Γ_1^+ ist nämlich eine Untergruppe von Γ_1' vom Index eins oder zwei; wäre daher Γ_1^+ endlich, so wäre auch Γ_1' im Widerspruch zu dem eben Gesagten endlich.

Im Falle eines zweidimensionalen Raums $\mathfrak{M}^{(1)}$ ist es nun unmittelbar einleuchtend, daß Γ_1^+ aus allen Rotationen um den Nullpunkt bestehen muß und somit $\mathfrak{M}^{(1)}$ sowie $\mathfrak{M}$ euklidisch sind, womit der Beweis von Satz 15 in diesem Falle schon erbracht ist. Es seien also im folgenden $\mathfrak{M}$ und $\mathfrak{M}^{(1)}$ als dreidimensional vorausgesetzt. Wir bezeichnen jetzt irgendeine unendliche Folge von (paarweise) verschiedenen Elementen aus Γ_1^+ mit $R_1, R_2, \ldots$, welche wir wegen der Kompaktheit von Γ_1^+ als gegen ein von allen R_i verschiedenes Element $R_0 \in \Gamma_1^+$ konvergierend annehmen dürfen. Dann konvergieren die Rotationen $R_i R_0^{-1}$ gegen die identische Abbildung I; sie sind alle von I verschieden und besitzen die Rotationsachsen a_i und die Rotationswinkel χ_i mit $\chi_i \neq 0, |\chi_i| \leqq \pi$ und $\lim_{i \to \infty} \chi_i = 0$ $(i = 1, 2, \ldots)$. Wir können weiter ohne Einschränkung der Allgemeinheit annehmen, daß die a_i gegen eine durch den Nullpunkt gehende Gerade a_0 konvergieren. Die in Γ_1^+ gelegene abgeschlossene Hülle der Rotationen $(R_i R_0^{-1})^\nu$ mit beliebigen ganzzahligen ν enthält nun sicher alle Rotationen um a_0, d.h. $\mathfrak{M}^{(1)}$ ist rotationssymmetrisch, und die Anwendung von Satz 14 zeigt in der Tat, daß $\mathfrak{M}$ euklidisch sein muß.

Wir erwähnen noch, *daß die Behauptung des eben bewiesenen Satzes für den Fall einer Dimension $n > 3$ falsch wird*, wie das Beispiel des Raums $\mathfrak{R}_{n-2} \times \mathfrak{M}_2$ zeigt, wobei $\mathfrak{R}_{n-2}$ ein $(n-2)$-dimensionaler euklidischer Raum und $\mathfrak{M}_2$ irgendeiner der in Satz 11 auftretenden Räume $\mathfrak{M}_2^{(j)}$ ist; w.z.b.w.

BEMERKUNGEN. 1. In Satz 14 und (im dreidimensionalen Fall) in Satz 15 ist die Voraussetzung einer vollhomogenen Norm wesentlich. Entsteht nämlich die Hyperfläche $\|x\| = 1$ eines n-dimensionalen Minkowski-Raums durch

alle Rotationen eines gleichschenkligen Dreiecks um diejenige Gerade d, welche seinen mit dem Nullpunkt zusammenfallenden Schwerpunkt und die zwischen seinen gleichlangen Seiten gelegene Ecke verbindet ($n>2$), so ist der auf diese Weise definierte Raum bezüglich d rotationssymmetrisch und selbstadjungiert, jedoch nicht euklidisch.

2. Ein selbstadjungierter Minkowski-Raum mit vollhomogener Norm der Dimension $n>2$ ist nicht schon durch die im Vergleich zu Satz 14 schwächere Bedingung als euklidisch gekennzeichnet, daß er mit einem Minkowski-Raum isomorph ist, welcher alle diejenigen (euklidischen) Bewegungen als Automorphismen gestattet, die einen durch den Nullpunkt gehenden, zweidimensionalen linearen Unterraum punktweise festlassen. Dies wird wiederum an dem vorigen Beispiel des Raums $\mathfrak{R}_{n-2}\times\mathfrak{M}_2$ deutlich.

3. Auch das letzte Korollar kann im Falle einer geraden Dimension $n>2$ nicht dahingehend abgeschwächt werden, daß man darin die Gerade d_1 durch einen den Nullpunkt enthaltenden, linearen Unterraum der Dimension $n'=n/2$ ersetzt. Um dies einzusehen, betrachten wir den Minkowski-Raum $\mathfrak{R}_n$ mit der Norm $\|x\|=\|\left((\xi_1^2+\cdots+\xi_{n'}^2)^{\frac{1}{2}},(\xi_{n'+1}^2+\cdots+\xi_n^2)^{\frac{1}{2}}\right)\|_{\mathfrak{R}_2}$, wobei $\|\ \|_{\mathfrak{R}_2}$ die den Bedingungen $\|(\xi_{n'},\xi_n)\|_{\mathfrak{R}_2}=\|(-\xi_{n'},\xi_n)\|_{\mathfrak{R}_2}=\|(\xi_{n'},-\xi_n)\|_{\mathfrak{R}_2}$ genügende Norm einer nichttrivialen selbstadjungierten Minkowski-Ebene $\mathfrak{R}_2$ mit dem zu $\mathfrak{A}=\begin{pmatrix}0&1\\1&0\end{pmatrix}$ gehörenden Isomorphismus A ist. Ein derartiges $\mathfrak{R}_2$ stellt z.B. die Minkowski-Ebene dar, bei welcher alle Punkte der Norm 1 ein reguläres Sechseck mit dem Nullpunkt als Schwerpunkt, dem Durchmesser $2(\cos\pi/6)^{-\frac{1}{2}}$ und einer auf einer der beiden Koordinatenachsen $\xi_n=0$ und $\xi_{n'}=0$ gelegenen Ecke bilden. Man weist auf Grund der, aus der Bedingung

$$\|(\xi_{n'},\xi_n)\|_{\mathfrak{R}_2}=\|(-\xi_{n'},\xi_n)\|_{\mathfrak{R}_2}=\|(\xi_{n'},-\xi_n)\|_{\mathfrak{R}_2}$$

folgenden Ungleichungen:

$$\|(\xi'_{n'},\xi'_n)\|_{\mathfrak{R}_2}\geqq\|(\xi_{n'},\xi_n)\|_{\mathfrak{R}_2}\quad\text{falls}\quad\xi'_i\geqq\xi_i\geqq 0\ (i=n',n)$$

leicht nach, daß die für $\mathfrak{R}_n$ definierte Norm tatsächlich alle Axiome eines linearen normierten Raums mit vollhomogener Norm erfüllt. Nun enthält $\mathfrak{R}_n$ den durch $\xi_1=\cdots=\xi_{n'-1}=\xi_{n'+1}=\cdots=\xi_{n-1}=0$ gegebenen nichttrivialen Minkowski-Raum $\mathfrak{R}_2$, ist also auch selbst nichttrivial. Weiter bewirken offensichtlich alle, den linearen Unterraum $\xi_{n'+1}=\cdots=\xi_n=0$ als Ganzes sowie den Nullpunkt festlassenden euklidischen Bewegungen Automorphismen von $\mathfrak{R}_n$. Für die Norm des Dualraums $\mathfrak{R}_n^*$ finden wir daher $\|f\|=\|\left((\alpha_1^2+\cdots+\alpha_{n'}^2)^{\frac{1}{2}},(\alpha_{n'+1}^2+\cdots+\alpha_n^2)^{\frac{1}{2}}\right)\|_{\mathfrak{R}_2^*}$, wenn wir mit $\|\ \|_{\mathfrak{R}_2^*}$ die Norm von $\mathfrak{R}_2^*$ bezeichnen. Jetzt stellt die Abbildung: $\xi_k=\alpha_{n'+k}$, $\xi_{n'+k}=\alpha_k\ (k=1,2,\ldots,n')$ einen Isomorphismus von $\mathfrak{R}_n^*$ auf $\mathfrak{R}_n$ dar, und damit ist $\mathfrak{R}_n$ als ein nichttrivialer, symmetrischer selbstadjungierter Minkowski-Raum mit den angegebenen Automorphismen erkannt. Übrigens ist noch die Isomorphie von $\mathfrak{R}_2$ mit dem Quotientenraum $\mathfrak{R}_n/\mathfrak{R}_{n-2}$[31]) bemerkenswert, wobei $\mathfrak{R}_{n-2}$ der durch $\xi_{n'}=\xi_n=0$ gegebene (abgeschlossene) Unterraum von $\mathfrak{R}_n$ sei.

[31]) Zur Definition eines Quotientenraums eines linearen normierten Raums mit vollhomogener Norm siehe [15], IX, § 3, Nr. 4, proposition 9.

Literatur

[1] JORDAN, P., and J. v. NEUMANN: On inner products in linear metric spaces. Ann. of Math. **36**, 719–723 (1935). — [2] LORCH, E. R.: On certain implications which characterize Hilbert space. Ann. of Math. **49**, 523–532 (1948). — [3] JAMES, R. C.: Inner products in normed linear spaces. Bull. Amer. Math. Soc. **53**, 559–566 (1947). — [4] DAY, M. M.: Some characterizations of inner-product spaces. Trans. Amer. Math. Soc. **62**, 320–337 (1947). — [5] RADON, J.: Über eine besondere Art ebener konvexer Kurven. Leipziger Ber. **68**, 123–128 (1916). — [6] LORCH, E. R.: The Cauchy-Schwarz inequality and self-adjoint spaces. Ann. of Math. **46**, 468–473 (1945). — [7] BOURBAKI, N.: Espaces vectoriels topologiques (livre V). Paris: Hermann 1953 u. 1955. — [8] LORCH, E. R.: Convexity and normed spaces. Acad. Serb. Sci. Publ. Inst. Math. **4**, 109–112 (1952). — [9] BANACH, S.: Théorie des opérations linéaires. Warschau 1932. — [10] NEUMANN, J. v.: Almost periodic functions in a group I. Trans. Amer. Math. Soc. **36**, 445–492 (1934). — [11] SCHMEIDLER, W.: Vorträge über Determinanten und Matrizen. Berlin: Akad. Verl. 1949. — [12] SCHREIER, O., u. E. SPERNER: Einführung in die analytische Geometrie und Algebra, Bd. II. Leipzig: B. G. Teubner 1935. — [13] KOKSMA, J. F.: Diophantische Approximationen. Erg. Math. 4. Berlin: Springer 1936. — [14] ALEXANDROFF, P., u. H. HOPF: Topologie. Berlin: Springer 1935. — [15] BOURBAKI, N.: Topologie générale (livre III). Paris: Hermann 1940–1949.

Freiburg i. B., Mathematisches Institut der Universität

(Eingegangen am 10. Januar 1959)

Barner, M., u. W. Degen
Math. Zeitschr. 71, 361–379 (1959)

Über sich schließende Figuren asymptotisch transformierter Kurven und W-transformierter Flächen

Von
Martin Barner und Wendelin Degen

Zwei punktweise aufeinander bezogene Kurven, die Asymptotenlinien der von ihnen erzeugten Regelfläche sind, nennt man asymptotisch Transformierte voneinander. Asymptotisch transformierte Kurven sind also dadurch charakterisiert, daß die Schmiegebenen in zusammengehörenden Punkten sich in der Verbindungsgeraden dieser Punkte schneiden. In enger Beziehung zu den asymptotisch transformierten Kurven stehen die W-transformierten Flächen: auf den beiden Brennflächen eines W-Strahlsystems bilden entsprechende Asymptotenlinien sowohl der einen wie der anderen Schar asymptotisch Transformierte voneinander.

Wir beschäftigen uns hier mit der Frage, ob man einparametrige Scharen von n-Ecken so angeben kann, daß je zwei benachbarte Eckpunktskurven asymptotisch Transformierte voneinander sind. Zu einer vorgegebenen Kurve kann man immer asymptotisch Transformierte konstruieren; es ist aber fraglich, ob sich die Figur, wenn man so fortfährt, zu einem n-Eck schließt. Daß sich dies nach vier Schritten so einrichten läßt, daß es also bewegte Vierecke dieser Art gibt, hat zuerst A. Terracini[1]) gezeigt. Für allgemeines n konstruieren wir hier Beispiele auf zwei verschiedene Arten: die eine Art besteht aus gewissen Systemen von Komplexkurven, deren Tangenten in einem gemeinsamen linearen Komplex liegen (es gehen $n-1$ willkürliche Funktionen einer Variabeln ein). Bei der andern Art handelt es sich um Bahnkurven einer Schiebung[2]) (es gehen vier willkürliche Funktionen ein). Indem man beide Möglichkeiten im Sinne der projektiven Kinematik kombiniert, d.h. das System aus Komplexkurven in die bewegte Basis einer Schiebung einsetzt, entsteht ein sich schließendes System von n Flächen, dessen Verbindungskongruenzen W-Strahlsysteme sind; die Komplexkurven und die Bahnkurven sind die Asymptotenlinien der verschiedenen Flächen.

In Sonderheit behandeln wir den Fall $n=5$. Hier sind nicht nur die genannten beiden Arten von Kurvensystemen die einzigen geschlossenen Figuren

[1]) Vgl. [1]; die Klammern [ ] weisen auf das Literaturverzeichnis am Schluß der Arbeit hin.

[2]) Die Theorie der Schiebungen wurde in den Arbeiten [2] entwickelt. Daraus benötigen wir hier nur einige wenige grundlegende Tatsachen, die am Anfang der dort genannten Abhandlung II hergeleitet worden sind. Diese Tatsachen werden hier zu Beginn des ersten Abschnittes so zusammengestellt, daß der Leser ohne Mühe ihre Richtigkeit bestätigen kann.

asymptotisch transformierter Kurven, sondern es sind auch die konstruierten Flächensysteme die einzigen sich mit der Periode fünf schließenden Figuren W-transformierter Flächen. Es gilt somit der Satz[3]): *Sich mit der Periode 5 schließende Figuren W-transformierter Flächen hängen von acht willkürlichen Funktionen einer Variablen ab. Die fünf Brennflächen sind (einsinnige) Komplexflächen; zusammengehörende Komplexkurven der verschiedenen Flächen gehören einem gemeinsamen linearen Komplex an. Die Figur läßt sich in einer speziellen projektiven Bewegung, in einer Schiebung, erzeugen. Die Parameterkurven sind die Asymptotenlinien der Flächen.*

Um die fünf Kurven bzw. Flächen gleichberechtigt und somit die ganze Figur symmetrisch behandeln zu können, verwenden wir eine analytische Methode, die sich schon bei der Behandlung geschlossener Laplace-Ketten bewährt hat[4]).

1. Bewegte Vierecke. Sind $\mathfrak{x}_0(t)$, $\mathfrak{x}_1(t)$, $\mathfrak{x}_2(t)$, $\mathfrak{x}_3(t)$ vier durch gleiche Parameterwerte t aufeinander bezogene Kurven und beschreiben $\mathfrak{x}_1$, $\mathfrak{x}_2$ Asymptotenlinien der durch sie bestimmten Regelfläche $(\mathfrak{x}_1, \mathfrak{x}_2)$ und gilt Entsprechendes für $\mathfrak{x}_2, \mathfrak{x}_3$ auf $(\mathfrak{x}_2, \mathfrak{x}_3)$, für $\mathfrak{x}_3, \mathfrak{x}_0$ auf $(\mathfrak{x}_3, \mathfrak{x}_0)$ und für $\mathfrak{x}_0, \mathfrak{x}_1$ auf $(\mathfrak{x}_0, \mathfrak{x}_1)$, so bestehen bei geeigneter Normierung der Vektoren $\mathfrak{x}_i$ die Ableitungsgleichungen

$$(1.1)\qquad \begin{cases} \mathfrak{x}_0' = a_0\mathfrak{x}_3 + c\mathfrak{x}_0 + a_3\mathfrak{x}_1, \\ \mathfrak{x}_1' = a_1\mathfrak{x}_0 - c\mathfrak{x}_1 - a_0\mathfrak{x}_2, \\ \mathfrak{x}_2' = a_2\mathfrak{x}_1 + c\mathfrak{x}_2 + a_1\mathfrak{x}_3, \\ \mathfrak{x}_3' = a_3\mathfrak{x}_2 - c\mathfrak{x}_3 - a_2\mathfrak{x}_0. \end{cases}$$

Darin ist die Normierung der Vektoren so gewählt, daß die Asymptotenlinien auf den genannten Regelflächen durch

$$(1.2)\qquad \mathfrak{x}_1 + u\mathfrak{x}_2, \quad \mathfrak{x}_2 + v\mathfrak{x}_3, \quad \mathfrak{x}_3 + w\mathfrak{x}_0 \qquad (u, v, w = \text{const})$$

beschrieben werden und ferner $(\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2, \mathfrak{x}_3) = \text{const}$ ist. Es zeigt sich, daß alle Ecken des Vierseits gleichberechtigt sind, daß also insbesondere auch die Asymptotenlinien auf $(\mathfrak{x}_0, \mathfrak{x}_1)$ durch $\mathfrak{x}_0 + s\mathfrak{x}_1$ mit $s = \text{const}$ gegeben sind.

Durch die so normierten Basisvektoren $\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2, \mathfrak{x}_3$, die den Ableitungsgleichungen (1.1) genügen, werden gewisse spezielle projektive Bewegungen definiert, die wir Schiebungen[5]) nannten. Bahnkurven in der Schiebung sind

---

[3]) Dieses Hauptergebnis der vorliegenden Arbeit wurde von beiden Verfassern unabhängig und von verschiedenen Ausgangspunkten aus gefunden. W. DEGEN untersuchte Flächensysteme aus fünf Flächen, die Brennflächen der Verbindungskongruenzen sind (sog. T_5-Figuren) und zeigte, daß die Figur, wenn die Kongruenzen W-Strahlsysteme sind, in einer allgemeinen Schiebung entsteht (Kolloquiumsvortrag gehalten am 22. 7. 58 in Kiel). Die T_5-Figuren wird W. DEGEN an anderer Stelle behandeln. — Die vorliegende Darstellung, die auf M. BARNER zurückgeht, bevorzugt den konstruktiven Aufbau, in dem auch allgemeinere Kurven- und Flächensysteme entstehen. Der Nachweis, daß so alle geschlossenen Figuren aus fünf W-Strahlkongruenzen erzeugt werden, mußte dann ebenfalls deduktiv geführt werden.

[4]) Vgl. M. BARNER [3].

[5]) Vgl. Fußnote 2).

die Kurven

$$\mathfrak{p} = p_0\mathfrak{x}_0 + p_1\mathfrak{x}_1 + p_2\mathfrak{x}_2 + p_3\mathfrak{x}_3 \qquad \text{mit } p_i = \text{const}. \tag{1.3}$$

Die Schiebungen sind unter den projektiven Bewegungen dadurch charakterisiert, daß die Tangenten an die Bahnkurven alle einem im bewegten System festen linearen Komplex angehören. Dieser Komplex wird in unserer Darstellung durch

$$\mathfrak{w} = (\mathfrak{x}_0, \mathfrak{x}_2) + (\mathfrak{x}_1, \mathfrak{x}_3) \tag{1.4}$$

repräsentiert; $\mathfrak{w}$ ist immer ein nichtausgearteter Komplex.

Wichtig ist nun für die weiteren Überlegungen, daß das Nullsystem, das zu dem ausgezeichneten Komplex $\mathfrak{w}$ gehört, jeden Punkt auf die Schmiegebene der durch ihn hindurchlaufenden Bahnkurve abbildet. Daraus folgt, daß eine im bewegten System feste Gerade g, die dem Komplex $\mathfrak{w}$ angehört, in der Schiebung eine Regelfläche erzeugt, auf der die Bahnkurven Asymptotenlinien sind. Denn die Schmiegebene an eine Bahnkurve enthält als Nullebene des Komplexes $\mathfrak{w}$ für jeden Wert des Parameters die Gerade g im Komplex $\mathfrak{w}$.

Wir heben hervor: *Sind* $\mathfrak{x}$ *und* $\mathfrak{y}$ *zwei feste Punkte des bewegten Systems einer Schiebung und liegt die Gerade* $(\mathfrak{x}, \mathfrak{y})$ *im ausgezeichneten Komplex* $\mathfrak{w}$, *so erzeugen die Punkte* $\mathfrak{x}$ *und* $\mathfrak{y}$ *in der Schiebung asymptotisch transformierte Kurven.*

Dieser Satz lehrt, wie sich Beispiele bewegter n-Ecke mit der geforderten Eigenschaft, daß benachbarte Eckpunktskurven asymptotisch transformierte voneinander sind, angeben lassen, falls wir die folgende elementargeometrische Aufgabe lösen können: Ein n-Eck $\overset{\circ}{\mathfrak{x}}_0, \overset{\circ}{\mathfrak{x}}_1, \ldots, \overset{\circ}{\mathfrak{x}}_{n-1}$ so zu bestimmen, daß die Seiten $(\overset{\circ}{\mathfrak{x}}_0, \overset{\circ}{\mathfrak{x}}_1), (\overset{\circ}{\mathfrak{x}}_1, \overset{\circ}{\mathfrak{x}}_2), \ldots, (\overset{\circ}{\mathfrak{x}}_{n-2}, \overset{\circ}{\mathfrak{x}}_{n-1})$ einem linearen Komplex angehören. Von $\overset{\circ}{\mathfrak{x}}_1$ ausgehend gelangt man zu $\overset{\circ}{\mathfrak{x}}_2$, indem man $\overset{\circ}{\mathfrak{x}}_2$ in der Nullebene von $\overset{\circ}{\mathfrak{x}}_1$ wählt usw. Es ist dann nur die Frage, ob sich die Figur schließen läßt. Seien etwa $\overset{\circ}{\mathfrak{x}}_1, \overset{\circ}{\mathfrak{x}}_2, \ldots, \overset{\circ}{\mathfrak{x}}_{n-1}$ schon konstruierte Punkte, dann ist $\overset{\circ}{\mathfrak{x}}_0$ so zu wählen, daß $(\overset{\circ}{\mathfrak{x}}_{n-1}, \overset{\circ}{\mathfrak{x}}_0)$ und $(\overset{\circ}{\mathfrak{x}}_0, \overset{\circ}{\mathfrak{x}}_1)$ Geraden im Komplex sind. Hierzu hat man $\overset{\circ}{\mathfrak{x}}_0$ auf der Schnittgeraden der Nullebene von $\overset{\circ}{\mathfrak{x}}_{n-1}$ und $\overset{\circ}{\mathfrak{x}}_1$ zu wählen und dies reicht auch hin. Störend würde sich nur auswirken, wenn diese beiden Nullebenen sich in der Geraden $(\overset{\circ}{\mathfrak{x}}_{n-1}, \overset{\circ}{\mathfrak{x}}_1)$ schneiden würden, so daß dann $\overset{\circ}{\mathfrak{x}}_0$ auf $(\overset{\circ}{\mathfrak{x}}_{n-1}, \overset{\circ}{\mathfrak{x}}_1)$ fiele, was wir natürlich nicht wünschen. In diesem Fall verwenden wir statt $\overset{\circ}{\mathfrak{x}}_{n-1}$ den Punkt $\overset{\circ}{\mathfrak{x}}^*_{n-1} = \overset{\circ}{\mathfrak{x}}_{n-1} + \lambda\overset{\circ}{\mathfrak{x}}_{n-2}$ mit $\lambda \neq 0$, dessen Nullebene dann sicher nicht durch $\overset{\circ}{\mathfrak{x}}_1$ geht. Die Nullebenen zweier verschiedener Punkte auf einer Geraden im Komplex sind ja verschiedene Ebenen durch die Verbindungsgerade dieser Punkte.

Wir sehen also, daß wir n-Ecke $\overset{\circ}{\mathfrak{x}}_0, \overset{\circ}{\mathfrak{x}}_1, \ldots, \overset{\circ}{\mathfrak{x}}_{n-1}$, dessen Seiten in einem Komplex liegen, ohne Schwierigkeiten konstruieren können. Bei dieser Konstruktion ist $\overset{\circ}{\mathfrak{x}}_1$ willkürlich, die Punkte $\overset{\circ}{\mathfrak{x}}_2, \ldots, \overset{\circ}{\mathfrak{x}}_{n-1}$, liegen der Reihe nach in einer vorgegebenen Ebene und $\overset{\circ}{\mathfrak{x}}_0$ fällt auf eine wohlbestimmte Gerade.

In einer Schiebung entsteht aus diesem n-Eck ein bewegtes n-Eck, wobei benachbarte Eckpunktskurven jeweils asymptotisch Transformierte sind. Da eine Schiebung durch die Ableitungsgleichungen (1.1) bestimmt ist und die Lösungen dieser Gleichungen bei willkürlicher Parameterverteilung von $c, a_0,$

a_1, a_2, a_3, und damit von vier willkürlichen Funktionen abhängen (wobei berücksichtigt ist, daß der Parameterverteilung keine geometrische Bedeutung zukommt) so folgt: *Es gibt für jedes n bewegte n-Ecke der Art, daß je zwei benachbarte Eckpunktskurven asymptotisch Transformierte voneinander sind. Solche bewegte n-Ecke hängen von mindestens vier willkürlichen Funktionen einer Veränderlichen (und von Konstanten) ab. Bei den angegebenen Beispielen entsteht die Figur aus einem n-Eck im bewegten System einer Schiebung.*

Es ist jedoch durchaus möglich und auch zu vermuten, daß es außer diesen in einer Schiebung erzeugbaren noch andere bewegte n-Ecke mit der genannten Eigenschaft gibt. Wir wenden uns deshalb der genaueren Behandlung des Falles $n=5$ zu.

2. Bewegte Fünfecke. Es seien $\mathfrak{x}_0(t), \ldots, \mathfrak{x}_4(t)$ die Eckpunktskurven eines bewegten Fünfecks des dreidimensionalen projektiven Raumes. Falls $\mathfrak{x}_1(t)$ Asymptotenlinie auf den Regelflächen $(\mathfrak{x}_0, \mathfrak{x}_1)$ und $(\mathfrak{x}_1, \mathfrak{x}_2)$ ist, wird die Schmiegebene an die Kurve $\mathfrak{x}_1(t)$ von $(\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2)$ aufgespannt. Also liegt auch die Tangente an die Kurve $\mathfrak{x}_1(t)$ in dieser Ebene, und es besteht die Ableitungsgleichung

$$\mathfrak{x}_1' = \alpha_1 \mathfrak{x}_0 + \gamma_1 \mathfrak{x}_1 + \beta_1 \mathfrak{x}_2. \tag{2.1}$$

Entsprechende Gleichungen gelten für die andern Eckpunktskurven. Diese Ableitungsgleichungen denken wir uns aus (2.1) durch zyklische Vertauschung der Indizes entstanden. Um anzudeuten, daß neben der Gl. (2.1) auch alle diejenigen bestehen, die durch zyklische Vertauschung der Indizes daraus hervorgehen, versehen wir diese Gleichung mit dem Zusatz (reihum). Wir schreiben also

$$\mathfrak{x}_1' = \alpha_1 \mathfrak{x}_0 + \gamma_1 \mathfrak{x}_1 + \beta_1 \mathfrak{x}_2 \qquad \text{(reihum)}. \tag{2.2}$$

Diese Gleichungen drücken aus, daß die Regelflächen $(\mathfrak{x}_0, \mathfrak{x}_1)$, $(\mathfrak{x}_1, \mathfrak{x}_2)$ sich längs der Kurve $\mathfrak{x}_1(t)$ berühren und daß Entsprechendes für die andern Eckpunktskurven gilt.

Zwischen den fünf Vektoren $\mathfrak{x}_0, \ldots, \mathfrak{x}_4$ besteht eine lineare Abhängigkeit

$$\varrho_0 \mathfrak{x}_0 + \varrho_1 \mathfrak{x}_1 + \varrho_2 \mathfrak{x}_2 + \varrho_3 \mathfrak{x}_3 + \varrho_4 \mathfrak{x}_4 = 0. \tag{2.3}$$

Um die Symmetrie zu erhalten, wollen wir nicht vier dieser Vektoren zu einer Basis auszeichnen; es sollen vielmehr alle gleichberechtigt behandelt werden. In (2.3) sind alle ϱ_i von Null verschieden; wäre etwa $\varrho_4 = 0$, so wären die Vektoren $\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2, \mathfrak{x}_3$ linear abhängig; unter den gemachten Voraussetzungen an unsere Figur würde das bedeuten, daß die Kurven $\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2, \mathfrak{x}_3$ Kurven einer gemeinsamen Ebene sind; es müßte dann auch $\mathfrak{x}_4$ in dieser Ebene verlaufen. — Wir wollen jedoch bewegte Fünfecke betrachten, die tatsächlich den dreidimensionalen Raum aufspannen.

Es ist also $\varrho_i \neq 0$ für jedes i, und wir können

$$\bar{\mathfrak{x}}_i = \varrho_i \mathfrak{x}_i \tag{2.4}$$

setzen und erhalten aus (2.3)

$$\bar{\mathfrak{x}}_0 + \bar{\mathfrak{x}}_1 + \bar{\mathfrak{x}}_2 + \bar{\mathfrak{x}}_3 + \bar{\mathfrak{x}}_4 = 0.$$

Wir denken uns die Normierung der Vektoren $\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_4$ von vornherein so eingerichtet; dann gilt also

$$\mathfrak{x}_0 + \mathfrak{x}_1 + \mathfrak{x}_2 + \mathfrak{x}_3 + \mathfrak{x}_4 = 0. \tag{2.5}$$

Diese Bedingung bleibt bei Umnormungen der Vektoren $\mathfrak{x}_i$ mit einem gemeinsamen Faktor erhalten.

Differenzieren wir (2.5), so folgt nach den Ableitungsgleichungen (2.2)

$$\sum_{(\text{reihum})} (\alpha_2 + \gamma_1 + \beta_0)\, \mathfrak{x}_1 = 0 \tag{2.6}$$

und da der Nullvektor nur auf die eine Art (2.5) aus den Vektoren $\mathfrak{x}_0, \ldots, \mathfrak{x}_4$ linear kombiniert werden kann, so folgt hieraus

$$\alpha_2 + \gamma_1 + \beta_0 = \varphi \qquad (\text{reihum}) \tag{2.7}$$

mit ein und demselben φ in allen Gleichungen. Ersetzen wir nach dieser Beziehung γ_1 in (2.2), so lauten die Ableitungsgleichungen

$$\mathfrak{x}_1' = \alpha_1 \mathfrak{x}_0 - (\alpha_2 + \beta_0 - \varphi)\, \mathfrak{x}_1 + \beta_1 \mathfrak{x}_2 \qquad (\text{reihum}). \tag{2.8}$$

Die aufgestellten Formeln (2.8) *mit* (2.5) *behandeln die Theorie der bewegten Fünfecke des dreidimensionalen projektiven Raumes mit der Eigenschaft, daß* $\mathfrak{x}_1$ *gemeinsame Kurve auf* $(\mathfrak{x}_0, \mathfrak{x}_1)$ *und auf* $(\mathfrak{x}_1, \mathfrak{x}_2)$ *ist und die beiden Regelflächen sich entlang dieser Kurve berühren und Entsprechendes reihum gilt. Die Normierung von* $\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2, \mathfrak{x}_3, \mathfrak{x}_4$ *ist bis auf eine gemeinsame Umnormungsfunktion festgelegt. Die Parameterverteilung ist willkürlich.*

Jetzt wenden wir uns dem Spezialfall zu, der uns interessiert; wir drücken aus, daß $\mathfrak{x}_1$ Asymptotenlinie auf $(\mathfrak{x}_0, \mathfrak{x}_1)$ und $(\mathfrak{x}_1, \mathfrak{x}_2)$ ist. Die Schmiegebene der Kurve $\mathfrak{x}_1$ ist dann die Ebene $(\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2)$ und es muß also

$$\mathfrak{x}_1'' \equiv 0 \pmod{\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2} \tag{2.9}$$

sein. Aus (2.8) folgt

$$\mathfrak{x}_1'' \equiv \alpha_0 \alpha_1 \mathfrak{x}_4 + \beta_1 \beta_2 \mathfrak{x}_3 \pmod{\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2} \tag{2.10}$$

und somit führt (2.9) wegen (2.5) auf $\alpha_0 \alpha_1 = \beta_1 \beta_2$. Da Entsprechendes auch für $\mathfrak{x}_2, \mathfrak{x}_3, \mathfrak{x}_4, \mathfrak{x}_0$ gelten soll, ist

$$\alpha_0 \alpha_1 = \beta_1 \beta_2 \qquad (\text{reihum}). \tag{2.11}$$

Der Ansatz $\alpha_0 = \varepsilon \beta_1$ führt nach (2.11) der Reihe nach auf

$$\alpha_1 = \frac{1}{\varepsilon} \beta_2, \quad \alpha_2 = \varepsilon \beta_3, \quad \alpha_3 = \frac{1}{\varepsilon} \beta_4, \quad \alpha_4 = \varepsilon \beta_0, \quad \alpha_0 = \frac{1}{\varepsilon} \beta_1. \tag{2.12}$$

Somit besteht für ε die Bedingung

$$\varepsilon^2 = 1. \tag{2.13}$$

Es kommen also für ε nur die beiden Werte $\varepsilon = 1$ und $\varepsilon = -1$ in Frage. Dabei wurde verwendet, daß keines der α_i, β_i verschwindet. Das Verschwinden einer dieser Größen widerspräche unserer Voraussetzung, daß $(\mathfrak{x}_1, \mathfrak{x}_2)$ (reihum) Regelflächen und keine Torsen sind.

Mit (2.12) nehmen unsere Ableitungsgleichungen die Gestalt

$$\mathfrak{x}_1' = \alpha_1 \mathfrak{x}_0 + (\varphi - \alpha_2 - \varepsilon \alpha_4)\, \mathfrak{x}_1 + \varepsilon \alpha_0 \mathfrak{x}_2 \quad \text{mit} \quad \varepsilon = \pm 1 \quad \text{(reihum)} \tag{2.14}$$

an. Eine Umnormung der Vektoren $\mathfrak{x}_i$ mit einem gemeinsamen Faktor, die noch frei ist, wirkt sich auf φ aus. Über diese Größe kann also noch verfügt werden (vgl. Abschnitt 7, Bemerkung 3). Da weiter die Parameterverteilung noch willkürlich ist, hängt unsere Figur tatsächlich von vier Funktionen einer Variablen (etwa bei $\varphi = 0$, $\alpha_0 = 1$ von $\alpha_1, \alpha_2, \alpha_3, \alpha_4$) ab. Wir heben hervor: *Die bewegten Fünfecke mit der Eigenschaft, daß benachbarte Ecken asymptotisch Transformierte beschreiben, zerfallen in zwei Scharen* ($\varepsilon = 1, \varepsilon = -1$). *Jede Schar hängt von vier willkürlichen Funktionen einer Veränderlichen ab.*

Wir werden im nächsten Abschnitt Unterscheidungsmerkmale beider Scharen angeben und ihre geometrischen Eigenschaften genauer beschreiben.

3. Die fünf Kurven $\mathfrak{x}_0, \ldots, \mathfrak{x}_4$ des projektiven Raumes bestimmen eindeutig eine projektive Bewegung, deren Bahnkurven durch

$$\mathfrak{p} = p_0 \mathfrak{x}_0 + p_1 \mathfrak{x}_1 + p_2 \mathfrak{x}_2 + p_3 \mathfrak{x}_3 + p_4 \mathfrak{x}_4 \quad \text{mit} \quad p_i = \text{const} \tag{3.1}$$

beschrieben werden. Da $\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2, \mathfrak{x}_3$ für jedes t eine Basis bilden, erfassen wir alle Bahnkurven der Bewegung bereits, wenn wir in (3.1) $p_4 = 0$ setzen. Wir untersuchen diese projektive Bewegung, d.h. die Gesamtheit ihrer Bahnkurven.

Aus Abschnitt 1 wissen wir, daß bei geeigneter Auswahl fünf Bahnkurven einer Schiebung ein Fünfeck asymptotisch transformierter Kurven bilden. Wir haben deshalb zuerst die Frage zu stellen: Wann ist die durch (2.14) bestimmte projektive Bewegung mit den Bahnkurven (3.1) eine Schiebung? Hierfür ist kennzeichnend, daß die Tangenten an die Bahnkurven alle einem im bewegten System festen linearen Komplex angehören.

Im Geradenraum verwenden wir als Basis zweckmäßigerweise die Bilder der fünf „Seiten" des bewegten Fünfecks

$$(\mathfrak{x}_0, \mathfrak{x}_1), \quad (\mathfrak{x}_1, \mathfrak{x}_2), \quad (\mathfrak{x}_2, \mathfrak{x}_3), \quad (\mathfrak{x}_3, \mathfrak{x}_4), \quad (\mathfrak{x}_4, \mathfrak{x}_0) \tag{3.2}$$

und den hierzu bezüglich der absoluten Quadrik des Geradenraumes polaren Punkt

$$\mathfrak{k} = \sum_{\text{(reihum)}} \{(\mathfrak{x}_0, \mathfrak{x}_1) + 2(\mathfrak{x}_0, \mathfrak{x}_2)\}. \tag{3.3}$$

$\mathfrak{k}$ stellt den linearen Komplex dar, dem für jedes t die Seiten des bewegten Fünfeckes angehören. Dieser Komplex ist eindeutig bestimmt, da die Seiten des Fünfecks nicht in einer linearen Kongruenz liegen. — Wegen der zyklischen Symmetrie der Darstellung von $\mathfrak{k}$ in (3.3) genügt es nachzuweisen, daß $(\mathfrak{x}_0, \mathfrak{x}_1)$

in dem Komplex mit der Darstellung (3.3) liegt. Diese Aussage führt auf die Relation

$$(\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2, \mathfrak{x}_3) + (\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_3, \mathfrak{x}_4) + 2(\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2, \mathfrak{x}_4) = 0,$$

die wegen (2.5) tatsächlich erfüllt ist.

Ist unsere projektive Bewegung eine Schiebung, so ist $\mathfrak{k}$ der in dieser Schiebung ausgezeichnete Komplex, denn das Nullsystem von $\mathfrak{k}$ bildet die Punkte $\mathfrak{x}_i$ auf die Schmiegebenen der von $\mathfrak{x}_i$ erzeugten Bahnkurven, das sind die Ebenen $(\mathfrak{x}_{i-1}, \mathfrak{x}_i, \mathfrak{x}_{i+1})$ $(i = 0, 1, 2, 3, 4;\ \mathfrak{x}_{-1} = \mathfrak{x}_4, \mathfrak{x}_5 = \mathfrak{x}_0)$, ab und ist das einzige Nullsystem mit dieser Eigenschaft. Ein solches Nullsystem gehört nämlich notwendig zu einem linearen Komplex, der die Seiten des Fünfecks enthält; durch diese Eigenschaft war aber $\mathfrak{k}$ gerade bestimmt. Wir fragen also danach, welche der Tangenten an Bahnkurven dem Komplex $\mathfrak{k}$ angehören und ob es möglich ist, daß alle Tangenten an Bahnkurven diese Eigenschaft haben. Drückt man aus, daß $(\mathfrak{p}, \mathfrak{p}')$ zu $\mathfrak{k}$ polar bezüglich der absoluten Quadrik des Geradenraumes ist (wobei $\mathfrak{p}$ in (3.1) definiert und $\mathfrak{p}'$ nach (2.14) zu berechnen ist), so folgt

$$(3.4) \qquad (1-\varepsilon) \sum_{(\text{reihum})} \{\alpha_0 \mathfrak{p}_0 \mathfrak{p}_1 + (-\alpha_0 - \alpha_1 + \alpha_3)\, \mathfrak{p}_0 \mathfrak{p}_2\} = 0.$$

Unsere projektive Bewegung ist eine Schiebung, wenn die Tangenten aller Bahnkurven dem Komplex $\mathfrak{k}$ angehören; dazu muß (3.4) identisch in den $\mathfrak{p}_i$ erfüllt sein. Dies führt auf $(1-\varepsilon) = 0$. Somit folgt: *Genau dann, wenn in* (2.14) $\varepsilon = 1$ *ist, sind die Eckpunktskurven des bewegten Fünfecks Bahnkurven einer Schiebung; diese Schiebung ist durch* (2.14) *mit* $\varepsilon = 1$ *bestimmt und die Gesamtheit ihrer Bahnkurven ist durch* (3.1) *gegeben.* $\mathfrak{k}$ *in* (3.3) *stellt den in der Schiebung ausgezeichneten Komplex dar.*

In Abschnitt 1 haben wir gezeigt, daß man in jeder Schiebung fünf Bahnkurven so auswählen kann, daß diese ein bewegtes Fünfeck bilden, dessen benachbarte Eckpunkte asymptotisch transformierte Kurven beschreiben. Deshalb gilt weiter: *Mit* $\varepsilon = 1$ *bestimmen die Ableitungsgleichungen* (2.14) *eine Schiebung, und jede Schiebung läßt sich unter Auszeichnung von fünf geeignet gewählten Bahnkurven so darstellen.* Man wird diese Darstellung einer Schiebung gegenüber (1.1) bevorzugen, wenn man sich für die hier betrachteten bewegten Fünfecke besonders interessiert.

Wir kommen zur weiteren Ausdeutung der Beziehung (3.4). Bei $\varepsilon = -1$ lehrt sie: *In der projektiven Bewegung* (2.14) *mit* $\varepsilon = -1$ *gehören für einen bestimmten Wert des Parameters* t *diejenigen Tangenten an Bahnkurven dem Komplex* $\mathfrak{k}$ *an, die von Punkten der Quadrik mit der Gleichung*

$$(3.5) \qquad \sum_{(\text{reihum})} \{\alpha_0 \mathfrak{p}_0 \mathfrak{p}_1 + (-\alpha_0 - \alpha_1 + \alpha_3)\, \mathfrak{p}_0 \mathfrak{p}_2\} = 0$$

auslaufen. Diese Quadrik ist im allgemeinen nicht fest im bewegten System.

Weiter zeigen wir: *Die Quadrik* (3.5) *berührt die Kurven* $\mathfrak{x}_0, \ldots, \mathfrak{x}_4$ *und ist durch diese Eigenschaft festgelegt.*

Zum Beweis fragen wir danach, wann eine in lokalen Koordinaten gegebene Quadrik

$$\langle \mathfrak{p}, \mathfrak{p} \rangle \equiv \sum_{i,k=0}^{4} a_{ik} p_i p_k = 0, \qquad a_{ik} = a_{ki} \tag{3.6}$$

die Eckpunktskurven $\mathfrak{x}_i(t)$ für einen bestimmten Wert des Parameters t berührt. Die Quadrik berührt die Kurve $\mathfrak{x}_1(t)$, wenn

$$f = f' = 0 \quad \text{mit} \quad f(t) = \langle \mathfrak{x}_1(t), \mathfrak{x}_1(t) \rangle \tag{3.7}$$

gilt. Nach den Ableitungsgleichungen (2.14) wird

$$f' \equiv \langle \mathfrak{x}_1, \alpha_1 \mathfrak{x}_0 + \varepsilon \alpha_0 \mathfrak{x}_2 \rangle = \alpha_1 \langle \mathfrak{x}_0, \mathfrak{x}_1 \rangle + \varepsilon \alpha_0 \langle \mathfrak{x}_1, \mathfrak{x}_2 \rangle \mod \langle \mathfrak{x}_1, \mathfrak{x}_1 \rangle$$

und (3.7) führt auf

$$\langle \mathfrak{x}_1, \mathfrak{x}_1 \rangle = 0, \qquad \alpha_1 \langle \mathfrak{x}_0, \mathfrak{x}_1 \rangle + \varepsilon \alpha_0 \langle \mathfrak{x}_1, \mathfrak{x}_0 \rangle = 0. \tag{3.8}$$

Wir wollen dem Vektor $\mathfrak{x}_i$ die Darstellung (3.1) mit $p_i = 1$, $p_k = 0$ für $k \neq i$ geben. Nach (3.6) ist dann $\langle \mathfrak{x}_i, \mathfrak{x}_k \rangle = a_{ik}$ und für die Koeffizienten der zu bestimmenden Quadrikgleichung bestehen also die Bedingungen

$$a_{11} = 0, \quad \alpha_1 a_{01} + \varepsilon \alpha_0 a_{12} = 0, \tag{3.9}$$

und zwar reihum, da nicht nur $\mathfrak{x}_1$, sondern alle Eckpunktskurven die Quadrik berühren sollten. Aus dem Ansatz $a_{01} = \varepsilon \alpha_0$ ergibt sich der Reihe nach

$$a_{12} = -\alpha_0, \quad a_{23} = \varepsilon \alpha_2, \quad a_{34} = -\alpha_3, \quad \alpha_{40} = \varepsilon \alpha_4, \quad a_{01} = -\alpha_0, \tag{3.10}$$

woraus als notwendige Bedingung für die Existenz einer solchen Quadrik $\varepsilon = -1$ folgt.

Mit $\varepsilon = -1$ lassen sich die Gln. (3.9) tatsächlich befriedigen und die dort genannten Koeffizienten a_{ik} sind bis auf einen gemeinsamen Faktor eindeutig bestimmt. Die weiteren a_{ik} findet man aus (3.8) wegen (2.5). Aus der ersten Gl. (3.8) folgt mit (2.5)

$$\langle \mathfrak{x}_1, \mathfrak{x}_0 \rangle + \langle \mathfrak{x}_1, \mathfrak{x}_2 \rangle + \langle \mathfrak{x}_1, \mathfrak{x}_3 \rangle + \langle \mathfrak{x}_1, \mathfrak{x}_4 \rangle = 0$$

und diese Beziehung muß auch reihum gelten. Diese fünf linearen Gleichungen für die Unbekannten $\langle \mathfrak{x}_0, \mathfrak{x}_2 \rangle$, $\langle \mathfrak{x}_1, \mathfrak{x}_3 \rangle$, $\langle \mathfrak{x}_2, \mathfrak{x}_4 \rangle$, $\langle \mathfrak{x}_3, \mathfrak{x}_0 \rangle$, $\langle \mathfrak{x}_4, \mathfrak{x}_1 \rangle$ liefern wegen $\langle \mathfrak{x}_0, \mathfrak{x}_2 \rangle = a_{02}$ und (3.10) als Lösung

$$a_{02} = -\alpha_0 - \alpha_1 + \alpha_3 \quad \text{(reihum)}. \tag{3.11}$$

Durch (3.9), (3.10) und (3.11) sind die Koeffizienten der Gleichung der gesuchten Quadrik bestimmt; wir gelangen, wie behauptet, wieder zu (3.5).

Über die ursprüngliche Behauptung hinaus haben wir damit gezeigt: *Die beiden Klassen von bewegten Fünfecken unterscheiden sich durch die Existenz einer lokal definierten Quadrik, die die fünf Eckpunktskurven in zusammengehörigen Punkten berührt; für $\varepsilon = -1$ existiert eine solche Quadrik, für $\varepsilon = 1$ dagegen nicht.*

Die wesentlichste Eigenschaft der bewegten Fünfecke mit $\varepsilon = -1$ finden wir aber aus den Ableitungsgleichungen im Geradenraum. Mittels (2.14) folgt

$$
(3.12)\quad \left\{\begin{aligned}
2(\mathfrak{x}_1, \mathfrak{x}_2)' = {} & (1+\varepsilon)\{-(\alpha_0+\alpha_1+\alpha_2+\alpha_3+\alpha_4)(\mathfrak{x}_1, \mathfrak{x}_2) + \alpha_1 \mathfrak{k}\} + \\
& + (1-\varepsilon)\{\alpha_1[(\mathfrak{x}_0, \mathfrak{x}_1) + (\mathfrak{x}_2, \mathfrak{x}_3) - (\mathfrak{x}_3, \mathfrak{x}_4) + (\mathfrak{x}_4, \mathfrak{x}_0)] + \\
& + (\alpha_0 - \alpha_2 - \alpha_3 + \alpha_4)(\mathfrak{x}_1, \mathfrak{x}_2)\} \qquad \text{(reihum)} \\
2\mathfrak{k}' = {} & (1+\varepsilon)\Big\{\sum_{\text{(reihum)}} A_1(\mathfrak{x}_1, \mathfrak{x}_2) - \Big(\sum_{i=0}^{4} \alpha_i\Big)\mathfrak{k}\Big\} \\
& \text{mit} \quad A_0 = \alpha_0 - \alpha_1 + \alpha_2 + \alpha_3 - \alpha_4 \qquad \text{(reihum)}.
\end{aligned}\right.
$$

Bei $\varepsilon = -1$ ist also der Komplex $\mathfrak{k}$ fest. Somit: *Bei $\varepsilon = -1$ gehören die Tangenten jeder der fünf Eckpunktskurven ein und demselben festen linearen Komplex an; diese Kurven sind also Komplexkurven in demselben Komplex $\mathfrak{k}$.*

Aus Komplexkurven desselben linearen Komplexes kann man aber, wie wir jetzt beweisen wollen, nicht nur bewegte Fünfecke, sondern auch n-Ecke so konstruieren, daß benachbarte Eckpunktskurven asymptotisch Transformierte voneinander sind.

Hierzu zeigen wir zweierlei: (I) Man kann alle zu einer gegebenen Komplexkurve asymptotisch transformierten Kurven desselben linearen Komplexes explizit bestimmen; diese hängen von einer willkürlichen Funktion einer Veränderlichen ab. (II) Zu zwei gegebenen Kurven $\mathfrak{x}(t)$ und $\mathfrak{z}(t)$ desselben linearen Komplexes, die nicht asymptotisch Transformierte voneinander sind, gibt es immer Kurven $\mathfrak{y}(t)$ mit der Eigenschaft, daß $\mathfrak{x}, \mathfrak{y}$ einerseits und auch $\mathfrak{y}, \mathfrak{z}$ andererseits asymptotisch Transformierte voneinander sind. Diese Kurven $\mathfrak{y}$ kann man bestimmen; sie hängen von einer Konstanten ab.

Nehmen wir beide Aussagen für den Augenblick als bewiesen an! Dann kann man nach (I) von einer Komplexkurve ausgehend, der Reihe nach $(n-1)$ der Eckpunktskurven des zu konstruierenden n-Ecks angeben. Dabei gehen für die Ausgangskomplexkurve eine und für jede weitere Kurve nach (I) eine weitere willkürliche Funktion einer Veränderlichen, insgesamt also $(n-1)$ Funktionen in die Figur ein. Seien etwa $\mathfrak{x}_3, \mathfrak{x}_4, \ldots, \mathfrak{x}_{n-1}, \mathfrak{x}_0, \mathfrak{x}_1$ diese Eckpunktskurven. Dann sind zwei Fälle zu unterscheiden. Entweder sind nun $\mathfrak{x}_1, \mathfrak{x}_3$ von selbst asymptotisch Transformierte voneinander geworden; unsere Figur schließt sich dann zu einem $(n-1)$-Eck; oder $\mathfrak{x}_1$ und $\mathfrak{x}_3$ sind nicht asymptotisch Transformierte voneinander und wir können nach (II) die Figur durch eine weitere Eckpunktskurve abschließen, wobei noch eine Konstante in die Betrachtung eingeht.

Im ersten Fall haben wir unser Ziel nicht erreicht, da wir ja ein n-Eck und nicht ein $(n-1)$-Eck konstruieren wollten. In diesem Fall verzichten wir auf den letzten Schritt, der uns die Kurve $\mathfrak{x}_1$ geliefert haben möge; die Figur $\mathfrak{x}_3, \mathfrak{x}_4, \ldots, \mathfrak{x}_{n-1}, \mathfrak{x}_0$ hängt von $n-2$ Funktionen ab. Wir gehen dann von $\mathfrak{x}_0$ und $\mathfrak{x}_3$ aus mit dem Ziel, diese beiden Komplexkurven durch zwei Kurven $\mathfrak{x}_1^*, \mathfrak{x}_2^*$ so zu verbinden, daß $\mathfrak{x}_0, \mathfrak{x}_1^*$ und $\mathfrak{x}_1^*, \mathfrak{x}_2^*$ und $\mathfrak{x}_2^*, \mathfrak{x}_3$ asymptotisch

Transformierte sind. Hierzu bestimmen wir nach (II) zunächst alle die Kurven, die zu $\mathfrak{x}_0$ und $\mathfrak{x}_3$ asymptotisch Transformierte sind ($\mathfrak{x}_1$ ist eine solche); diese hängen von einer Konstanten ab. Dann wählen wir nach (I) eine solche asymptotisch Transformierte $\mathfrak{x}_1^*$ zu $\mathfrak{x}_0$, die nicht auch zu $\mathfrak{x}_3$ asymptotisch Transformierte ist. Dann können wir wieder nach (II) die Figur durch die Kurve $\mathfrak{x}_2^*$ zu einem n-Eck abschließen.

Wir haben die Beweise für (I) und (II) nachzutragen und dabei darauf zu achten, daß die zuletzt verwendete Auswahl möglich ist.

Zu (I). Es sei eine Komplexkurve $\mathfrak{x}(t)$ des Komplexes $\mathfrak{k}$ gegeben und es seien $(\mathfrak{x}, \mathfrak{x}') = \mathfrak{g}$ die Tangenten an die Komplexkurve. Wir wollen eine beliebige Regelfläche auswählen, auf der die Ausgangskurve Asymptotenlinie ist.. Auf dieser Regelfläche gibt es dann im allgemeinen genau eine weitere Asymptotenlinie, deren Tangenten in dem Komplex $\mathfrak{k}$ liegen. Wir werden diese Tangenten explizit bestimmen.

Da die Ausgangskurve Asymptotenlinie der Regelfläche sein soll, liegt die Erzeugende $\mathfrak{r}$ der Regelfläche $\mathfrak{r}(t)$ in der Schmiegebene und geht durch den Kurvenpunkt; es ist also $\mathfrak{r} = \alpha(\mathfrak{x}, \mathfrak{x}') + \beta(\mathfrak{x}, \mathfrak{x}'')$. Im Geradenraum ist mit $\mathfrak{g} = (\mathfrak{x}, \mathfrak{x}')$ auch $\mathfrak{g}' = (\mathfrak{x}, \mathfrak{x}'')$ und also $\mathfrak{r}$ und, weil $\mathfrak{k}$ fest ist, damit auch $\mathfrak{r}'$, $\mathfrak{r}''$ zu $\mathfrak{k}$ bezüglich der absoluten Quadrik polar. Durch $(\mathfrak{r}, \mathfrak{r}', \mathfrak{r}'')$ wird die eine Erzeugendenschar der Schmiegquadrik der Regelfläche bestimmt, und die dazu total polare Ebene der Dimension zwei des Geradenraumes bestimmt die andere Geradenschar dieser Quadrik; es sind dies die Tangenten an die Asymptotenlinien der Regelfläche längs einer Erzeugenden. Unter diesen Geraden interessiert uns nur die eine, $\mathfrak{h}$, die außer $\mathfrak{g}$ in dem Komplex $\mathfrak{k}$ liegt. Im Geradenraum ist das polare Gebilde zu $(\mathfrak{r}, \mathfrak{r}', \mathfrak{r}'', \mathfrak{k})$ eine Gerade, die die absolute Quadrik außer in $\mathfrak{g}$ noch in einem davon verschiedenen Punkte $\mathfrak{h}$ trifft. So entsteht längs der Regelfläche eine einparametrige Geradenschar $\mathfrak{h}(t)$ und es bleibt zu zeigen, daß die Geraden dieser Schar die Tangenten einer Asymptotenlinie von $\mathfrak{r}(t)$ sind. Dafür ist kennzeichnend, daß $\mathfrak{h}, \mathfrak{h}', \mathfrak{r}$ linear abhängig sind und die durch $\mathfrak{h}, \mathfrak{h}', \mathfrak{r}$ bestimmte Gerade auf der absoluten Quadrik liegt. Dies folgt aber daraus, daß die voneinander verschiedenen Punkte $\mathfrak{h}, \mathfrak{h}', \mathfrak{r}$ zueinander polar sind und alle drei auf der polaren Geraden von $(\mathfrak{h}, \mathfrak{r}, \mathfrak{r}', \mathfrak{k})$ liegen.

Zu (II). Es seien $\mathfrak{x}(t)$ und $\mathfrak{z}(t)$ zwei Kurven in einem gemeinsamen linearen Komplex. Diese Kurven seien nicht asymptotisch Transformierte voneinander. Soll $\mathfrak{y}(t)$ asymptotisch Transformierte sowohl zu $\mathfrak{x}$ wie auch zu $\mathfrak{z}$ sein, so muß $\mathfrak{y}$ für jeden Wert des Parameters t in den beiden Schmiegebenen an die Kurven $\mathfrak{x}(t)$ und $\mathfrak{z}(t)$ liegen. Diese Schmiegebenen sind aber die Nullebenen von $\mathfrak{x}$ und $\mathfrak{z}$ in dem Nullsystem, das zu $\mathfrak{k}$ gehört. $\mathfrak{y}$ liegt auf der Schnittgeraden dieser Nullebenen; diese Schnittgerade $\mathfrak{s}$ ist windschief zu der Geraden $(\mathfrak{x}, \mathfrak{z})$.

Wenn $\mathfrak{y}$ auf $\mathfrak{s}$ asymptotisch Transformierte von $\mathfrak{x}$ und $\mathfrak{z}$ ist, so geht die Schmiegebene an die Kurve $\mathfrak{y}$ durch $\mathfrak{x}$ und $\mathfrak{z}$; die Tangente an die Kurve $\mathfrak{y}$ trifft also die Gerade $(\mathfrak{x}, \mathfrak{z})$.

Es sei $\mathfrak{y}$ eine Kurve der Kurvenschar auf $\mathfrak{s}(t)$, deren Tangenten für jedes t die Gerade $(\mathfrak{x}, \mathfrak{z})$ treffen. $\mathfrak{y}$ ist dann, wie wir zeigen wollen, asymptotisch

Transformierte zu $\mathfrak{x}$ und $\mathfrak{z}$. Die Nullebene von $\mathfrak{y}$ im Nullsystem von $\mathfrak{k}$ ist nämlich die Ebene $(\mathfrak{y}, \mathfrak{x}, \mathfrak{z})$. Also liegen die Tangenten der Kurve $\mathfrak{y}$ in dem Komplex $\mathfrak{k}$, und die genannte Nullebene ist auch die Schmiegebene der Kurve $\mathfrak{y}$. Dies war zu beweisen.

Verwenden wir zwei der Kurven $\mathfrak{y}_1, \mathfrak{y}_2$ auf $\mathfrak{F}(t)$, deren Tangenten $(\mathfrak{x}, \mathfrak{z})$ treffen, so erhalten wir in $\mathfrak{x}, \mathfrak{y}_1, \mathfrak{z}, \mathfrak{y}_2$ ein bewegtes Viereck asymptotisch transformierter Kurven, wie wir es in Abschnitt 1 betrachtet haben. Alle vier Eckpunktskurven liegen jetzt in einem festen Komplex $\mathfrak{k}$. Verwenden wir die Ableitungsgleichungen (1.1) aus Abschnitt 1, so muß der feste Komplex $\mathfrak{k}$ mit geeignetem λ und μ durch $\lambda(\mathfrak{x}_0, \mathfrak{x}_2) + \mu(\mathfrak{x}_1, \mathfrak{x}_3)$ dargestellt sein. Nach (1.1) ist dieser Komplex genau dann fest, wenn $\lambda = -\mu$ ist und in den Gln. (1.1) $c = 0$ gilt.

Die Kurven auf den Diagonalen $(\mathfrak{x}_0, \mathfrak{x}_3)$ unseres Vierecks, die zu $\mathfrak{x}_1$ und $\mathfrak{x}_2$ asymptotisch transformiert sind, werden durch $\mathfrak{x}_0 + \tau\mathfrak{x}_3$ mit $\tau = \text{const}$ gegeben. Wählen wir nach (I) eine Regelfläche, deren Erzeugende durch $\mathfrak{x}_1$ gehen und in der Schmiegebene von $\mathfrak{x}_1$ liegen — diese Regelfläche wird dann dargestellt durch $(\mathfrak{x}_1, \mathfrak{x}_0 + \sigma\mathfrak{x}_3)$ —, so gibt es auf dieser Regelfläche eine zu $\mathfrak{x}_1$ asymptotisch transformierte Kurve des Komplexes $\mathfrak{k}$, die genau dann nicht zu $\mathfrak{x}_2$ asymptotisch transformiert ist, wenn σ nicht konstant ist. Diese Tatsache hatten wir oben zur Konstruktion der n-Ecke aus asymptotisch transformierten Komplexkurven benötigt.

Wir fassen das Ergebnis zusammen: *Es gibt für jedes $n \geqq 4$ bewegte n-Ecke mit der Eigenschaft, daß benachbarte Eckpunktskurven asymptotisch Transformierte voneinander sind und alle einem festen linearen Komplex angehören.* Die Ausgangskomplexkurve hängt von einer Funktion einer Veränderlichen ab, ebenso die der Reihe nach konstruierten Kurven $\mathfrak{x}_2, \mathfrak{x}_3, \ldots, \mathfrak{x}_{n-1}$. $\mathfrak{x}_0$ ist dann bis auf eine Konstante bestimmt. *Die Figur der so bewegten n-Ecke hängt also von $n-1$ willkürlichen Funktionen einer Veränderlichen ab.*

4. Als Folgerung: Sich schließende Figuren von W-Strahlsystemen. Wir haben auf zwei verschiedene Arten bewegte n-Ecke der Eigenschaft konstruiert, daß benachbarte Eckpunktskurven asymptotisch Transformierte sind: im ersten Fall (Abschnitt 1) sind die Eckpunktskurven Bahnkurven einer Schiebung, im zweiten Fall (Abschnitt 3) sind die Eckpunktskurven Komplexkurven ein und desselben linearen Komplexes. Ferner sahen wir (Abschnitt 2), daß im Falle $n = 5$ nur diese beiden Fälle möglich sind.

Verdienen diese Figuren auch an sich Beachtung, so kommt ihnen darüber hinaus besonderes geometrisches Interesse dadurch zu, daß sich die beiden projektiven Bewegungsvorgänge zusammensetzen lassen, so daß ein geschlossenes System von Flächen entsteht, wobei jede Fläche Brennfläche der anliegenden Verbindungskongruenzen ist. Die Zusammensetzung der beiden Bewegungsvorgänge erfolgt, indem wir das in Abschnitt 3 konstruierte System von Komplexkurven mit den sie verbindenden Regelflächen, auf denen die Komplexkurven Asymptotenlinien sind, in das bewegte System einer Schiebung einsetzen, und zwar so, daß diese Figur in dem in der Schiebung ausgezeichneten Komplex liegt, welcher im bewegten System fest ist. Dann entsteht

aus jeder der Komplexkurven $\overset{\circ}{\mathfrak{x}}_0(s), \ldots, \overset{\circ}{\mathfrak{x}}_{n-1}(s)$ in der Schiebung eine Komplexfläche $\mathfrak{x}_0(s,t), \ldots, \mathfrak{x}_{n-1}(s,t)$. Die Tangentenebene an die Fläche $\mathfrak{x}_1$ ist die Ebene $(\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2)$, denn diese Ebene ist die Nullebene von $\mathfrak{x}_1$ im Nullsystem des ausgezeichneten Komplexes. Diese Ebene ist als Nullebene Schmiegebene der Bahnkurve durch $\mathfrak{x}_1$. Dies heißt, die Parameterkurven sind Asymptotenlinien jeder der so konstruierten Flächen $\mathfrak{x}_0, \ldots, \mathfrak{x}_{n-1}$. Die Gerade $(\mathfrak{x}_1, \mathfrak{x}_2)$ liegt in der Tangentenebene jeder der beiden Flächen $\mathfrak{x}_1, \mathfrak{x}_2$; die Regelfläche $\left(\overset{\circ}{\mathfrak{x}}_1(s), \overset{\circ}{\mathfrak{x}}_2(s)\right)$ des bewegten Systems erzeugt also in der Schiebung eine Geradenkongruenz, deren Brennflächen die Flächen $\mathfrak{x}_1$ und $\mathfrak{x}_2$ sind. Diese Geradenkongruenz ist ein W-Strahlsystem, denn die Asymptotenlinien entsprechen sich als die Parameterkurven auf beiden Brennflächen.

Wir fassen zusammen: *Setzt man das bewegte n-Eck aus Komplexkurven $\overset{\circ}{\mathfrak{x}}_0(s), \overset{\circ}{\mathfrak{x}}_1(s), \ldots, \overset{\circ}{\mathfrak{x}}_{n-1}(s)$ in das bewegte System einer Schiebung ein, so entsteht eine geschlossene Figur aus n W-Strahlsystemen, deren Brennflächen Komplexflächen der Art sind, daß zusammengehörende Komplexkurven der verschiedenen Flächen in demselben linearen Komplex liegen. Die sich entsprechenden Asymptotenlinien auf den Brennflächen sind die Komplexkurven einerseits und die Bahnkurven in der Schiebung andererseits.* Da eine Schiebung von vier Funktionen und das geschlossene System aus n Komplexkurven von $n-1$ Funktionen einer Veränderlichen abhängt, so folgt weiter: *Die geschlossene Figur der n W-Strahlsysteme dieser Art hängt von $n+3$ willkürlichen Funktionen einer Veränderlichen ab.*

Man weiß[6]), daß ein W-Strahlsystem mit seinen Brennflächen sich immer dann aus den Komplexkurven als Asymptotenlinien einer Komplexregelfläche in einer Schiebung erzeugen läßt, wenn die Brennflächen des Systems Komplexflächen sind und zusammengehörende Komplexkurven in demselben linearen Komplex liegen. Da weiter die hierbei benötigte Schiebung durch jede der Brennflächen voll bestimmt wird, gilt eine Umkehrung in der folgenden Form: *Hat man eine (geschlossene) Figur aus W-Strahlsystemen, deren Brennflächen Komplexflächen sind, und liegen zusammengehörende Komplexkurven der verschiedenen Flächen in demselben linearen Komplex, so wird die Figur in einer Schiebung, wie oben beschrieben, aus einem System von Komplexkurven gewonnen.*

Für $n=5$ läßt sich aber viel mehr beweisen. Hier sind die genannten Figuren die einzigen, die ein geschlossenes System von fünf W-Kongruenzen bilden.

5. Sich schließende Figuren von fünf W-Strahlsystemen. In diesem Abschnitt beweisen wir den Satz: *Hat man im projektiven dreidimensionalen Raum fünf Flächen $\mathfrak{x}_0(s,t), \mathfrak{x}_1(s,t), \ldots, \mathfrak{x}_4(s,t)$, so daß $\mathfrak{x}_i, \mathfrak{x}_{i+1}$ die Brennflächen der Verbindungskongruenz $(\mathfrak{x}_i, \mathfrak{x}_{i+1})$ sind, und sind diese Kongruenzen W-Strahlsysteme, so sind die Brennflächen $\mathfrak{x}_0, \ldots, \mathfrak{x}_4$ Komplexflächen und zusammengehörende Komplexkurven liegen in demselben linearen Komplex. Die Figur entsteht in einer allgemeinen Schiebung aus einem bewegten Fünfeck, dessen benachbarte*

[6]) Vgl. M. BARNER [2], III, S. 320.

Eckpunktskurven asymptotisch Transformierte und außerdem Komplexkurven desselben linearen Komplexes sind. Die eine Schar der sich auf den Brennflächen entsprechenden Asymptotenlinien sind die Komplexkurven, die andere wird aus Bahnkurven in der Schiebung gebildet.

Zum Beweis würde es nach dem zuletzt ausgesprochenen Satz in Abschnitt 4 genügen, zu zeigen, daß zusammengehörende Asymptotenlinien der einen Schar der verschiedenen Flächen jeweils in einem gemeinsamen linearen Komplex liegen. Wir werden aber darüber hinaus auch noch zeigen, daß die Kurven der andern Schar von Asymptotenlinien Bahnkurven einer Schiebung sind; damit ist dann die volle Aussage ohne Rückgriff auf die Theorie der Schiebungen hier bewiesen.

Diese Aussagen folgen fast unmittelbar aus unseren vorbereitenden Überlegungen und Ergebnissen der Abschnitte 1—3. Da es sich bei unserer Figur um fünf W-Strahlsysteme handelt, entsprechen sich die Asymptotenlinien auf den fünf Flächen. Durchlaufen wir zusammengehörende Asymptotenlinien der Flächen $\mathfrak{x}_1$ und $\mathfrak{x}_2$, so sind diese Kurven auch auf der durch sie bestimmten Regelfläche $(\mathfrak{x}_1, \mathfrak{x}_2)$ Asymptotenlinien. Daraus folgt, daß zusammengehörende Asymptotenlinien in der Figur reihum asymptotisch Transformierte voneinander sind. Dies gilt für beide Scharen von Asymptotenlinien.

Daß nun die eine Sorte Asymptotenlinien Bahnkurven einer Schiebung, die andere dagegen Komplexkurven eines gemeinsamen Komplexes sind, zeigt man am besten analytisch. Sind die Parameterkurven $s = \text{const}$, $t = \text{const}$ Asymptotenlinien, so bestehen, da sie reihum asymptotisch Transformierte voneinander sind, nach (2.14) die Ableitungsgleichungen

$$(5.1)\quad \begin{cases} \mathfrak{x}_{1t} = \alpha_1 \mathfrak{x}_0 - (\alpha_2 + \varepsilon\,\alpha_4 - \varphi)\,\mathfrak{x}_1 + \varepsilon\,\alpha_0 \mathfrak{x}_2 & \text{(reihum)} \\ \mathfrak{x}_{1s} = \bar{\alpha}_1 \mathfrak{x}_0 - (\bar{\alpha}_2 + \bar{\varepsilon}\,\bar{\alpha}_4 - \bar{\varphi})\,\mathfrak{x}_1 + \bar{\varepsilon}\,\bar{\alpha}_0 \mathfrak{x}_2 & \text{(reihum).} \end{cases}$$

Hier ist zu zeigen, daß $\varepsilon\bar{\varepsilon} = -1$ ist, denn dies bedeutet nach Abschnitt 3, daß die Asymptotenlinien der einen Schar Komplexkurven, die der andern Schar Bahnkurven einer Schiebung sind. Die Beziehung $\varepsilon\bar{\varepsilon} = -1$ folgt aus den Integrierbarkeitsbedingungen

$$(5.2)\qquad \mathfrak{x}_{1ts} - \mathfrak{x}_{1st} = 0 \qquad \text{(reihum).}$$

Nach (5.1) und wegen (2.5) ergibt sich hieraus bei Vergleich der Faktoren von $\mathfrak{x}_4$ und $\mathfrak{x}_3$

$$(5.3)\qquad (1 + \varepsilon\,\bar{\varepsilon})\,(\alpha_1 \bar{\alpha}_3 - \bar{\alpha}_1 \alpha_0) = 0,$$

also entweder, wie gewünscht, $\varepsilon\bar{\varepsilon} = -1$, oder $\alpha_1\bar{\alpha}_0 - \bar{\alpha}_1\alpha_0$ (reihum). Im letzteren Fall entartet aber die Figur, denn $\alpha_1\bar{\alpha}_0 - \bar{\alpha}_1\alpha_0 = 0$ bedeutet nach (5.1), daß sich die t- und s-Kurven berühren. Es bestehen in der Figur keine zwei unabhängigen Richtungen: die Figur ist in ein System von Kurven entartet, wie wir es oben betrachtet haben. Nach Ausschluß dieses Falles bleibt also $\varepsilon\bar{\varepsilon} = -1$, und wir können ohne Einschränkung der Allgemeinheit $\varepsilon = 1$, $\bar{\varepsilon} = -1$ setzen.

Wir wissen dann nach Abschnitt 3, daß für jeden Wert von $t = t_0$ die s-Asymptotenlinien in einem gemeinsamen linearen Komplex liegen. Nach der Theorie der Schiebungen könnte man daraus alles Weitere folgern.

Man kann aber auch direkt beweisen, daß die Figur aus den eben genannten Komplexkurven in einer projektiven Bewegung entsteht. Diese projektive Bewegung muß dann nach Abschnitt 2 eine Schiebung sein. Hierzu ist nur zu zeigen, daß die Koeffizienten $\bar{\alpha}_i$ in den Ableitungsgleichungen (5.1) von t unabhängig sind. Dies ergibt sich aus den weiteren Integrierbarkeitsbedingungen. Sie lauten nach Umformung

$$(5.4)\quad \begin{cases} \bar{\alpha}_{1t} = 0 \qquad \text{(reihum)} \\ \alpha_{1s} = \alpha_1(\bar{\alpha}_0 - \bar{\alpha}_2 - \bar{\alpha}_3 + \bar{\alpha}_4) - \bar{\alpha}_1(\alpha_0 - \alpha_2 + \alpha_3 - \alpha_4) \qquad \text{(reihum)} \\ \bar{\varphi}_t - \varphi_s = \alpha_0\bar{\alpha}_1 - \bar{\alpha}_0\alpha_1 + \alpha_1\bar{\alpha}_2 - \bar{\alpha}_1\alpha_2 + \alpha_2\bar{\alpha}_3 - \bar{\alpha}_2\alpha_3 + \alpha_3\bar{\alpha}_4 - \bar{\alpha}_3\alpha_4 + \alpha_4\bar{\alpha}_0 - \bar{\alpha}_4\alpha_0. \end{cases}$$

Die Gleichungen der ersten Zeile drücken aus, daß, wie behauptet, die $\bar{\alpha}_i$ Funktionen von t allein sind.

Damit ist der zu Beginn dieses Abschnitts ausgesprochene Satz voll bewiesen. Auf weitere geometrische Einzelheiten, die sich aus der Theorie der Komplexflächen als Schiebflächen ergeben, gehen wir hier nicht ein.

6. Spezielle bewegte Fünfecke, Laplace-Ketten. Wir kehren zu den bewegten Fünfecken zurück, bei denen benachbarte Eckpunktskurven asymptotisch Transformierte voneinander sind. Zusammengehörende Asymptotenlinien der sich schließenden Figuren von W-Strahlsystemen bilden, wie wir sahen, solche so bewegten Fünfecke.

Durch die Figur und die zyklische Symmetrie sind die Strahlbüschel durch die Eckpunkte $\mathfrak{x}_0, \ldots, \mathfrak{x}_4$ in den zugehörigen Ebenen $(\mathfrak{x}_4, \mathfrak{x}_0, \mathfrak{x}_1), \ldots, (\mathfrak{x}_3, \mathfrak{x}_4, \mathfrak{x}_0)$ projektiv aufeinander bezogen, wenn man erstens $(\mathfrak{x}_1, \mathfrak{x}_2)$ (reihum), zweitens $(\mathfrak{x}_0, \mathfrak{x}_1)$ (reihum), drittens die Tangenten $(\mathfrak{x}_1, \alpha_1\mathfrak{x}_0 + \varepsilon\alpha_0\mathfrak{x}_2)$ (reihum) einander zuordnet. Es entsprechen sich dann die Geraden

$$(6.1)\qquad \alpha_1(\mathfrak{x}_0, \mathfrak{x}_1) - \varepsilon\mu\,\alpha_0(\mathfrak{x}_1, \mathfrak{x}_2) \qquad \text{(reihum)}$$

mit ein und demselben μ in allen Gleichungen.

Die Geraden (6.1) gehören sämtlich dem Komplex $\mathfrak{k}$ an. Können, so wollen wir fragen, zusammengehörende Gerade in einer linearen Kongruenz liegen? Die fünf Punkte des Geradenraumes

$$(6.2)\qquad \alpha_1(\mathfrak{x}_0, \mathfrak{x}_1) - \varepsilon\mu\,\alpha_0(\mathfrak{x}_1, \mathfrak{x}_2) \qquad \text{(reihum)}$$

sind genau dann linear abhängig, wenn

$$(6.3)\qquad (\varepsilon\mu)^5 = 1 \qquad \text{also} \qquad \varepsilon\mu = 1$$

gilt. Daraus folgt: *Zu der Figur eines bewegten Fünfecks gehört lokal eine lineare Kongruenz, der bei $\varepsilon = 1$, d.h. im Falle einer Schiebung, die Tangenten an die Eckpunktskurven und bei $\varepsilon = -1$, d.h. im Falle der Komplexkurven, die zu den Tangenten bezüglich der Seiten harmonischen Geraden angehören.*

Im Geradenraum ist diese Kongruenz in beiden Fällen durch $\mathfrak{k}$ und

$$\mathfrak{l} = \sum_{(\text{reihum})} A_1(\mathfrak{x}_1, \mathfrak{x}_2) \tag{6.4}$$

bestimmt. Hierin ist

$$A_0 = \alpha_0 - \alpha_1 + \alpha_2 + \alpha_3 - \alpha_4 \qquad (\text{reihum}). \tag{6.5}$$

Bei $\varepsilon = 1$, d.h. im Falle einer Schiebung ist diese Kongruenz vielfach ausgezeichnet. Ihr gehören nicht nur die Tangenten an die Eckpunktskurven, sondern die Tangenten aller Bahnkurven an. Diese Kongruenz ist bei $\varepsilon = 1$ im Geradenraum nach (3.12) und (6.4) durch $(\mathfrak{k}, \mathfrak{l})$ bestimmt und enthält somit die Gesamtheit der Geraden, deren Bildpunkte im Geradenraum momentane Fixpunkte sind. Die Punkte der Leitlinien der Kongruenz sind die momentanen Fixpunkte der Schiebung im dreidimensionalen Raum.

Kann diese mit der Figur invariant verbundene Kongruenz fest sein? Dies bedeutet analytisch

$$\mathfrak{k}' \equiv 0, \quad \mathfrak{l}' \equiv 0 \pmod{\mathfrak{k}, \mathfrak{l}}. \tag{6.6}$$

Die erste dieser Bedingungen ist für $\varepsilon = 1$ nach (3.12) und (6.4) erfüllt; für $\varepsilon = -1$ folgt sie daraus, daß $\mathfrak{k}$ fest ist. Die zweite führt bei $\varepsilon = 1$ auf

$$A_i' = \psi A_i \qquad (i = 0, 1, 2, 3, 4) \tag{6.7}$$

und bei $\varepsilon = -1$ auf

$$A_1' + A_1(\alpha_0 - \alpha_2 - \alpha_3 + \alpha_4) - \alpha_2 A_2 + \alpha_3 A_3 - \alpha_4 A_4 + \alpha_0 A_0 = \psi A_1 \quad (\text{reihum}). \tag{6.8}$$

Wir gehen auf die Größen α_i zurück und verwenden, daß nach (6.5)

$$A_0 + A_1 = 2\alpha_3 \qquad (\text{reihum}) \tag{6.9}$$

gilt. Damit folgt aus (6.7) unmittelbar

$$\alpha_i' = \psi \alpha_i \tag{6.10}$$

und aus (6.8) nach einiger Rechnung

$$\alpha_3' = \alpha_3(\alpha_0 - \alpha_1) + \chi \alpha_3 \qquad (\text{reihum}). \tag{6.11}$$

Die α_i sind alle von Null verschieden vorausgesetzt. Bei einer Parametertransformation auf unserm Kurvensystem

$$t^* = f(t), \qquad \frac{dt^*}{dt} = \varphi(t) \tag{6.12}$$

gilt

$$\alpha_i^* = \varphi^{-1} \alpha_i. \tag{6.13}$$

Wir können deshalb durch geeignete Wahl der Parameterverteilung erreichen, daß

$$\alpha_0 \alpha_1 \alpha_2 \alpha_3 \alpha_4 = \text{const} \tag{6.14}$$

ist. Differentiation dieser Beziehung lehrt, daß bei dieser Parameterverteilung die oben eingeführten Funktionen ψ, χ in (6.10) bzw. (6.11) verschwinden. Es gilt also bei geeigneter Parameterverteilung in den Spezialfällen:

(6.15) $\qquad \alpha_i = \text{const} \qquad (\varepsilon = 1,\ i = 0, 1, 2, 3, 4)$

bzw.

(6.16) $\qquad \alpha_3' = \alpha_3(\alpha_0 - \alpha_1) \qquad (\varepsilon = -1)$ (reihum).

Wir heben hervor: *Ist die mit unserm Kurvensystem lokal invariant verbundene Geradenkongruenz längs des Kurvensystems fest, so gilt bei geeigneter Parameterverteilung bei* $\varepsilon = 1$ (6.15), *bei* $\varepsilon = -1$ (6.16).

Die Schiebungen mit dieser Eigenschaft sind leicht zu beschreiben. Da in diesen die Tangenten an die Bahnkurven der festen linearen Kongruenz angehören, entarten alle Bahnkurven in gerade Linien. Dagegen ist die spezielle Konfiguration von Komplexkurven ($\varepsilon = -1$ und (6.16)) nicht so leicht zu erfassen. Hier besteht (vgl. Abschnitt 7, Bemerkung 4) eine sehr enge Koppelung zwischen den einzelnen Kurven des Kurvensystems und jede Kurve hat, da die Konfiguration nur noch von Konstanten abhängt, speziellen Charakter.

Gehen wir dann wieder zu den geschlossenen Figuren von fünf W-Strahlsystemen über, so werden hier auch diese Flächensysteme besonders auffallen, bei denen zusammengehörende Asymptotenlinien sowohl der einen wie auch der andern Schar die spezielle Eigenschaft haben, daß die zugehörige lineare Kongruenz fest ist. Diese beiden so dem Flächensystem zugeordneten linearen Kongruenzen brauchen natürlich im allgemeinen nicht zusammenzufallen. Es zeigt sich aber, daß die Forderung des Zusammenfallens beider Kongruenzen deren Festsein bereits nach sich zieht, so daß wir dann genau auf unsern betrachteten Spezialfall geführt werden.

Nach (6.4) in Verbindung mit dem Abschnitt 5 ist für das Zusammenfallen der beiden linearen Kongruenzen

(6.17) $\qquad A_i = \varrho \bar{A}_i \qquad (i = 0, 1, 2, 3, 4)$

kennzeichnend. Hieraus folgt wieder unmittelbar

(6.18) $\qquad \alpha_i = \varrho \bar{\alpha}_i \qquad (i = 0, 1, 2, 3, 4).$

Indem man nun zwar die Parameterkurven auf unserm Flächensystem festhält, aber innerhalb der Parameterkurven zu einer andern Parameterverteilung übergeht, kann man hier ϱ zu 1 machen. Bei dieser Parameterverteilung gilt also dann in unserm Spezialfall

(6.19) $\qquad \alpha_i = \bar{\alpha}_i \qquad (i = 0, 1, 2, 3, 4).$

Die Funktionen $\alpha_i(s) = \bar{\alpha}_i(s)$ genügen dann nach (5.4) den Ableitungsgleichungen (6.16). (Durch Umnormung der Vektoren $\mathfrak{x}_0, \ldots, \mathfrak{x}_4$ kann man nun $\varphi = \bar{\varphi} = 0$ erreichen.)

Geometrisch bedeutet das Zusammenfallen der beiden linearen Kongruenzen, daß die t-Tangente im Flächenpunkt $\mathfrak{x}_1$ (reihum) derselben linearen

Kongruenz angehört wie die vierte Harmonische zur s-Tangente bezüglich der Geraden $(\mathfrak{x}_0, \mathfrak{x}_1)$, $(\mathfrak{x}_1, \mathfrak{x}_2)$; da es jedoch durch einen Raumpunkt nur eine einzige Kongruenzgerade gibt, so heißt dies, daß die s- und t-Richtungen konjugiert sind bezüglich den Richtungen $(\mathfrak{x}_0, \mathfrak{x}_1)$, $(\mathfrak{x}_1, \mathfrak{x}_2)$. Dies heißt aber genau, daß aufeinanderfolgende W-Strahlsysteme einer Laplace-Kette angehören. Unsere geschlossene Figur der fünf W-Strahlsysteme ist also jetzt eine geschlossene Laplace-Kette, die man, da ja jede ihrer Kongruenzen ein W-Strahlsystem ist, auch R-Kette nennt.

Mit (6.19) *stellt unsere Figur eine geschlossene R-Kette der Periode fünf dar. Die geschlossenen R-Ketten der Periode* 5 *hängen also noch von vier Konstanten ab.* Wir haben diese Figur bereits an anderer Stelle[7]) erwähnt; sie wird in einer Arbeit von W. DEGEN genauer untersucht werden.

7. *Ergänzende Bemerkungen — Hinweise auf offene Fragen*

1. Gibt es zu zwei punktweise aufeinander bezogenen Kurven $\mathfrak{x}(t)$, $\mathfrak{y}(t)$ eine dritte Kurve $\mathfrak{z}(t)$ mit der Eigenschaft, daß $\mathfrak{x}, \mathfrak{z}$ und $\mathfrak{y}, \mathfrak{z}$ asymptotisch Transformierte sind, so nennt man die Kurven des Paares asymptotisch Transformierte zweiter Stufe voneinander. In einer geschlossenen Figur asymptotisch transformierter Kurven mit vier oder fünf Kurven sind je zwei beliebig herausgegriffene Kurven asymptotisch Transformierte erster oder zweiter Stufe. Schon bei einer Figur aus sechs Kurven ist dies nicht mehr der Fall.

Es ist zu vermuten, daß zwei willkürliche, punktweise aufeinander bezogene Kurven (in dieser Sprechweise) immer asymptotisch Transformierte dritter Stufe sind. Ein allgemeines Kurvenpaar hängt nämlich von fünf willkürlichen Funktionen ab; die zu einer vorgegebenen Kurve asymptotisch Transformierten hängen von einer willkürlichen Funktion ab. Die zu einer Kurve in dritter Stufe asymptotisch Transformierten hängen deshalb von drei willkürlichen Funktionen und die ganze Figur einschließlich der Ausgangskurve also ebenfalls von fünf Funktionen ab.

In erster bzw. zweiter Stufe asymptotisch Transformierte können nicht auch in zweiter bzw. erster Stufe transformiert sein, weil sonst die Figur in eine Ebene entarten würde. Asymptotisch Transformierte erster oder zweiter Stufe sind aber immer auch asymptotisch Transformierte dritter Stufe. Asymptotisch Transformierte erster bzw. zweiter Stufe lassen sich nämlich immer als Bahnkurven einer Schiebung auffassen.

Wenn sich beweisen läßt, daß jedes Kurvenpaar in erster, zweiter oder dritter Stufe asymptotisch transformiert ist, so lassen sich leicht geschlossene Figuren asymptotisch transformierter Kurven mit mehr als fünf Eckpunktskurven konstruieren.

2. Eine Kurve auf einer Regelfläche des fünfdimensionalen Raumes heißt quasi-asymptotisch, falls die Schmiegebene der Dimension drei dieser Kurve in dem Raum liegt, der aufgespannt wird von den Tangenten an die Regelfläche

[7]) Vgl. M. BARNER [3].

längs einer Erzeugenden und von der Schmiegebene der Dimension zwei der betrachteten Kurve. Durch jeden Punkt der Erzeugenden der Regelfläche gibt es, wie im R_3, eine Quasi-Asymptotenlinie. Zwei Quasi-Asymptotenlinien einer Regelfläche des R_5 nennen wir wieder asymptotisch Transformierte voneinander.

Man betrachte im R_5 geschlossene Figuren aus sechs bzw. sieben Regelflächen der Art, daß benachbarte Regelflächen sich in der gemeinsamen Kurve berühren und diese auf beiden Regelflächen Quasi-Asymptotenlinie ist. Man übertrage die Ergebnisse der Abschnitte 1, 2 und 3 auf diese Figuren. Die bewegten Siebenecke dieser Art zerfallen nicht (wie die Fünfecke des dreidimensionalen Raumes) in zwei Klassen.

Entsprechende Überlegungen kann man in den Räumen mit beliebiger ungerader Dimension anstellen. Die dieser Art bewegten $(n+2)$-Ecke des R_n zerfallen in zwei Klassen, wenn $n = 4k-1$ ist, dagegen nicht, wenn $n = 4k+1$ ist (k ist eine beliebige natürliche Zahl).

3. Die Ableitungsgleichungen (2.8) lassen sich durch eine Umnormung aller Eckpunktsvektoren mit einem gemeinsamen Faktor

$$\bar{\mathfrak{x}}_i = \varrho\,\mathfrak{x}_i \qquad (\ln\varrho)' = -\varphi \tag{7.1}$$

dahingehend vereinfachen, daß man die Funktion φ zum Verschwinden bringt. Nach der Umnormung (wir lassen den Querstrich gleich wieder weg) gilt

$$\mathfrak{x}_1' = \alpha_1\mathfrak{x}_0 - (\alpha_2+\beta_0)\,\mathfrak{x}_1 + \beta_1\mathfrak{x}_2 \qquad \text{(reihum)}. \tag{7.2}$$

Wir haben in Abschnitt 2 und 3 im Hinblick auf die Folgerungen des Abschnitts 5 darauf verzichtet, diese Vereinfachung durchzuführen. Man kann nämlich bei den Flächensystemen durch eine gemeinsame Umnormung nicht mehr erreichen, daß die dort auftretenden Funktionen φ und $\bar{\varphi}$ verschwinden. Dies ist dann und nur dann möglich, wenn

$$\sum_{i=0}^{4} \alpha_i\bar{\alpha}_{i+1} - \bar{\alpha}_i\alpha_{i+1} = 0. \tag{7.3}$$

(Vergleiche die Integrierbarkeitsbedingungen (5.4).)

4. Auf Grund der Lieschen Geraden-Kugeltransformation ist eine Komplexkurve äquivalent den Krümmungskreisen einer Kurve der im Sinne der Lie-Geometrie erweiterten euklidischen Ebene. Zwei asymptotisch transformierten Komplexkurven entsprechen dann zwei Kurven der euklidischen Ebene, die Hüllkurven einer Kreisschar sind. In diesem Modell sieht man, daß es zu zwei vorgegebenen Kurven der euklidischen Ebene eine dritte Kurve der Art gibt, daß die erste und dritte Kurve sowie die zweite und dritte Kurve die Hüllkurve je einer Kreisschar sind. Damit ist das Ergebnis des Beweises in Abschnitt 3 geometrisch illustriert.

Die spezielle Konfiguration aus Komplexkurven mit (6.17) ist dann einem Kreissystem äquivalent, wobei die Berührkreise jeder der fünf Kurven eng gekoppelt sind. Diese Berührkreise sind harmonisch zueinander bezüglich des Krümmungskreises und eines Kreises aus einem fest vorgegebenen linearen

Kreissystem. Ohne Rücksicht auf die Realitätsverhältnisse (und unter Außerachtlassung eines Spezialfalles) kann man das feste Kreissystem mit den Punkten der euklidischen Ebene identifizieren; man kommt so zu der Figur von fünf Kurven der euklidischen Ebene der Art, daß benachbarte Kurven Hüllkurven einer Kreisschar sind. Jede Kurve ist so Hüllkurve von zwei Kreisscharen, deren Kreise in dem Berührbüschel den Krümmungskreis der Kurve vom Kurvenpunkt für jeden Parameterwert harmonisch trennen.

5. Man kann geschlossene R-Ketten, deren Netzflächen Kongruenzregelflächen sind, mit einer beliebigen Periodenzahl konstruieren, wenn man in einer Schiebung mit $\alpha_i = \text{const}$ (s. (6.15)) ein System von Komplexkurven $\mathfrak{x}_0, \mathfrak{x}_1, \cdots, \mathfrak{x}_{n-1}$ so vorgibt, daß die Tangente an die Komplexkurve $\mathfrak{x}_1$ von der Geraden durch $\mathfrak{x}_1$ in der ausgezeichneten linearen Kongruenz durch $(\mathfrak{x}_0, \mathfrak{x}_1)$ und $(\mathfrak{x}_1, \mathfrak{x}_2)$ harmonisch getrennt wird, und Entsprechendes reihum gilt. In der Lie-geometrischen Deutung kommt man zu Kurvensystemen der euklidischen Ebene wie sie am Ende der Bemerkung 4 beschrieben sind, wobei statt von fünf nun von n Kurven zu sprechen wäre.

6. Nach Abschnitt 4 kann man sich schließende Figuren aus W-Strahlsystemen mit beliebiger Anzahl n der Brennflächen konstruieren. Für $n = 5$ sind dies nach Abschnitt 5 auch die einzig möglichen. Für $n = 4$ gibt es jedoch noch andere. Auch für $n = 6$ (man vgl. die Bemerkung 1) sind noch andere Lösungen zu erwarten. Kann man das Problem etwa dadurch symmetrisch behandeln, daß man für die sechs Punkte $\mathfrak{x}_0, \mathfrak{x}_1, \mathfrak{x}_2, \mathfrak{x}_3, \mathfrak{x}_4, \mathfrak{x}_5$ des dreidimensionalen projektiven Raumes zwei lineare Bedingungen als gegeben voraussetzt?

7. Mit einem bewegten Fünfeck asymptotisch transformierter Kurven ist lokal invariant eine Quadrik verbunden. In Abhängigkeit des Parameters t entsteht eine Quadrikschar, die von vier willkürlichen Funktionen abhängt. Durch welche Eigenschaften ist diese Quadrikschar unter den allgemeinen einparametrigen Quadrikscharen, die von neun Funktionen abhängen, gekennzeichnet? Was läßt sich über die charakteristischen Kurven vierter Ordnung, die Schnittkurven benachbarter Quadriken, aussagen?

Literatur

[1] Terracini, A.: Sulle superficie con una sistema di assintotiche in complessi lineari. Torino Atti **59**, 441–461 (1924). – [2] Barner, M.: Zur projektiven Differentialgeometrie der Komplexflächen. I. Komplexflächen als Schiebflächen. Math. Annalen **126**, 119–137 (1953). – II. Konstruktion und integrallose Darstellung spezieller Schiebungen. Math. Annalen **126**, 418–446 (1953). – III. Schiebungen als Abrollvorgänge. Singularitätenkurven auf Komplexflächen. Invariante Gebilde. Math. Annalen **129**, 304–322 (1955). – [3] Barner, M.: Über geschlossene Laplace-Ketten. Archiv d. Math. **9**, 366–377 (1958). – [4] Degen, W.: Geschlossene Laplace-Ketten der Periode fünf im dreidimensionalen Projektiven Raum und ihre n-dimensionalen Verallgemeinerungen. Dissertation, Freiburg i. Br. 1958.

Karlsruhe, Lehrstuhl für Darstellende Geometrie der Technischen Hochschule
Freiburg i. Br., Mathematisches Institut der Universität

(Eingegangen am 17. Januar 1959)

COHN, P. M.
Math. Zeitschr. 71, 380–398 (1959)

On the Free Product of Associative Rings

By
P. M. COHN

Introduction

A ring R is said to be the free product of a family of rings R_ι $(\iota \in I)$ if the R_ι are subrings of R such that (i) R is generated by the "factors" R_ι, (ii) every relation in R is a consequence of relations in the factors R_ι, and (iii) different factors intersect in the same subring Λ of R. The object of this paper is to give conditions under which the free product of associative rings exists.

In order to describe these conditions we require the notions of "flatness" and "purity". Given any associative ring Λ with a unit-element, let U be a right Λ-module with a submodule U' and let V be a left Λ-module; then we may form the (not necessarily exact) sequence

$$0 \to U' \otimes_\Lambda V \to U \otimes_\Lambda V. \tag{0.1}$$

If, for a given V, (0.1) is exact for all U, U', then V is said to be (left) *flat*; if U and U' are given and (0.1) is exact for all V, then U' is said to be (right) *pure* in U. The definitions of "right flat" and "left pure" are similar (cf. § 2). For brevity we shall call a ring R with Λ as subring (with the same unit-element as R) a Λ-ring (cf. § 1). It is then clear that R may be considered as a Λ-bimodule in a natural way. Now we have the following conditions for the existence of free products.

I. Given any family of Λ-rings R_ι $(\iota \in I)$, if each R_ι/Λ is left flat, then the free product R of the R_ι exists and R/Λ is again left flat. A similar assertion holds with "left" replaced by "right" throughout.

II. Given a Λ-ring R, if for every Λ-ring S the free product of R and S exists, then Λ is both right pure and left pure in R.

The sufficient condition I is proved in § 4 (Theorem 4.5) and the necessary condition II in § 5 (Theorem 5.2 Cor.). These sections also contain some further conditions and a result on the structure of the free product (Theorem 4.6). It seems likely that more precise conditions for the existence of the free product could be established in terms of the Tor groups of the factors, but a criterion of this sort might be rather complicated, and hence more difficult to apply.

A number of special cases of I and II are perhaps worth mentioning. From I and a theorem of M. HARADA and M. AUSLANDER [1] we deduce that the free product of Λ-rings always exists if Λ is a regular ring. This includes the elementary (and probably well known) case where Λ is a division ring.

From II, together with certain conditions for purity given in § 5, examples are readily constructed of rings whose free product does not exist. Such rings may be obtained e.g. by taking Λ to be the ring of polynomials in one variable over a field.

In § 3 we prove some general theorems on free products and universal products of rings which have for the most part analogues in group theory[1]). Indeed many could be formulated in the general theory of algebraic systems and are no doubt well known, but no convenient reference seems to be available[2]).

In the final section (§ 6) we consider the case where Λ is without a unit-element, but we confine our attention throughout to the case were all the intersections of pairs of factors coincide, corresponding to the case of the product of groups with a single amalgamated subgroup. This case was first treated by O. SCHREIER [*14*], who showed that such a free product always exists. More recently it has been shown by Z. E. DIDIDSE [*8*] that in the case of non-associative linear algebras[3]) the free product with one amalgamated subalgebra always exists, in contrast to the situation in the associative case, treated here.

1. Λ-rings

Let Λ be a ring[4]) with a unit-element, denoted by 1. A ring R is said to be a *Λ-ring*, if it is a unital Λ-bimodule such that[5])

$$(1.1)\quad \lambda(xy) = (\lambda x)\, y, \quad (x\lambda)\, y = x(\lambda y), \quad (xy)\,\lambda = x(y\lambda) \quad (x, y \in R, \lambda \in \Lambda).$$

The Λ-rings form a class of algebras[6]), the operations being the ring-operations together with the unary operations of multiplying by the elements of Λ (on the right or left). In this sense we understand by a subring of a Λ-ring a subring admitting Λ, and homomorphisms, isomorphisms etc. of Λ-rings are understood to be Λ-homomorphisms, Λ-isomorphisms etc.

If the Λ-ring R has a unit-element e, say, and if $\lambda e = e\lambda$ for all $\lambda \in \Lambda$, then it is easily seen, by (1.1), that the mapping $\varphi: \lambda \to e\lambda$ is a (ring-)homomorphism of Λ into R. If moreover Λ acts faithfully on R, then φ is a monomorphism, and the embedding so defined allows us to identify Λ with a subring of R. The unit-elements of Λ and R then coincide and both may be denoted by 1. Conversely, if R is any ring with a unit-element 1, and Λ any subring of R which contains 1, then R may be regarded as a Λ-ring in a natural way.

---

[1]) Cf. e.g. B. H. NEUMANN [*12*].

[2]) R. SIKORSKI [*15*] treats the problem of the existence of the free product of general algebraic systems, but his work does not overlap ours in any way.

[3]) Cf. also KUROŠ [*11*].

[4]) All rings and algebras are understood to be associative.

[5]) In the more usual case of a Λ-algebra the operator domain is essentially commutative (BOURBAKI [*3*] p. 101). The above definition is wider in that it allows any ring with 1 as unit-element and containing Λ as a subring to be regarded as a Λ-ring.

[6]) Cf. BIRKHOFF [*2*] p. vii.

The Λ-ring R will be said to be a Λ-ring *in the strict sense*, if R has a unit-element and Λ acts faithfully on R (qua set of right multiplications), to distinguish it from a Λ-ring *in the wide sense*, which also has a unit-element, but where Λ need not act faithfully. We shall only be dealing with Λ-rings with a unit-element, and they will nearly always be Λ-rings in the strict sense; we therefore drop the epithet in this case, so that from now on "Λ-ring" stands for "Λ-ring in the strict sense". The above remarks show that a Λ-ring is then essentially no more than a ring with Λ as a subring. The advantage of the terminology lies in the fact that a Λ-ring contains Λ in a canonical fashion. We note further that in requiring our rings to have a unit-element 1, we admit as subrings only the rings containing 1 (and hence containing Λ), and a homomorphism from R to S must map the 1 of R on the 1 of S. In other words, we regard 1 as a 0-ary operator on our class of algebras. From this point of view Λ itself may be regarded as the free Λ-ring on 0 free generators. Similarly a Λ-ring in the wide sense is essentially a ring containing a homomorphic image of Λ as a subring in a canonical fashion.

2. Flat modules and pure submodules

For the rest of this paper, except § 6, Λ will be a fixed (but arbitrary) ring with 1. Unless the contrary is stated, all tensor products are taken over Λ, and we shall write Tor for Tor_1^Λ (cf. CARTAN & EILENBERG [5] p. 107).

A left Λ-module V is said to be *flat* if for every right Λ-module X, $\mathrm{Tor}(X, V) = 0$. Clearly this is equivalent to the definition given in the introduction (cf. [5] p. 122). A corresponding definition applies for right Λ-modules. It is easily verified that (i) every free module is flat, (ii) every direct summand of a flat module is flat, and (iii) if V has a submodule V' such that both V' and V/V' are flat, then so is V. The last property follows from the exact homology sequence of the tensor product ([5] p. 39). The homology sequence can also be used to show that a submodule of a flat module is not necessarily flat. A further property which will be used is given in

LEMMA 2.1. *Let S be any ring with 1. If U is an (S, Λ)-bimodule*[7]) *which is flat, qua left S-module, and V is a flat left Λ-module, then $U \otimes_\Lambda V$, qua left S-module, is flat.*

For let

$$0 \to X' \to X \tag{2.1}$$

be an exact sequence of right S-modules. Since U is a flat left S-module, the sequence

$$0 \to X' \otimes_S U \to X \otimes_S U$$

is exact, and since V is a flat left Λ-module, the sequence

$$0 \to X' \otimes_S U \otimes_\Lambda V \to X \otimes_S U \otimes_\Lambda V$$

is exact. The result now follows from the associativity of the tensor product.

[7]) Cf. CARTAN & EILENBERG [5] p. 22.

Given a right Λ-module U with a submodule U', we defined U' to be *pure* in U in the introduction, if the sequence

$$0 \to U' \otimes V \to U \otimes V \tag{2.2}$$

is exact for all left Λ-modules V. Equivalently we may define U' to be pure in U if the natural homomorphism $\mathrm{Tor}(U, V) \to \mathrm{Tor}(U/U', V)$ is an epimorphism. In particular, U' is pure in U if U/U' is flat (but not conversely). More generally, by considering the exact sequence

$$\mathrm{Tor}(U', V) \to \mathrm{Tor}(U, V) \to \mathrm{Tor}(U/U', V)$$

for the case where U' is flat, we obtain

THEOREM 2.2. *Let U be a right Λ-module with submodule U' and suppose that* (i) *U' is flat and* (ii) *U' is pure in U. Then for every left Λ-module V, the natural homomorphism*

$$\mathrm{Tor}(U, V) \to \mathrm{Tor}(U/U', V)$$

is an isomorphism. In particular, under the given conditions U is flat if and only if U/U' is flat.

In order to obtain a criterion for purity we require a lemma.

LEMMA 2.3. *If V is a finitely presented left Λ-module, with generators $v_1^*, \ldots, v_n^*$ and relations $\sum \lambda_{ij} v_j^* = 0$ $(i = 1, \ldots, m)$, and if for some set of elements $u_1, \ldots, u_n$ in a right Λ-module U, we have*

$$\sum u_j \otimes v_j^* = 0 \qquad \text{in } U \otimes V,$$

then $u_j = \sum a_i \lambda_{ij}$, for some elements $a_i \in U$.

Proof. Let F be the free left Λ-module on $v_1, \ldots, v_n$ as free generators, and let R be the submodule of F generated by $\sum \lambda_{ij} v_j$ $(i = 1, \ldots, m)$. The mapping $v_j \to v_j^*$ may be extended to a homomorphism π of F into V; from the definition of V it follows that π is an epimorphism with kernel R. We therefore have the exact sequence

$$0 \to R \xrightarrow{\iota} F \xrightarrow{\pi} V \to 0,$$

which induces the exact sequence

$$U \otimes R \xrightarrow{\iota'} U \otimes F \xrightarrow{\pi'} U \otimes V \to 0. \tag{2.3}$$

We are given $(\sum u_j \otimes v_j)\pi' = 0$, hence by the exactness of (2.3),

$$\sum u_j \otimes v_j = (\sum a_i \otimes \lambda_{ij} v_j)\, \iota',$$

for some elements $a_i \in U$. Now ι' is the homomorphism induced by the inclusion mapping $\iota: R \to F$, and F is free, hence equating the coefficients of v_j, we obtain $u_j = \sum a_i \lambda_{ij}$, as we wished to show.

We can now state a necessary and sufficient condition for a Λ-submodule to be pure.

THEOREM 2.4. *Let U be a right Λ-module. Then a submodule U' of U is pure in U if and only if, for any finite sets of elements $u_i \in U$, $a_j \in U'$, $\lambda_{ij} \in \Lambda$ $(i = 1, \ldots, m; j = 1, \ldots, n)$, the relations*

$$a_j = \sum u_i \lambda_{ij} \tag{2.4}$$

imply the existence of elements $u_i' \in U'$ such that $a_j = \sum u_i' \lambda_{ij}$. Moreover, it is sufficient for the purity of U' in U that the condition hold when the u_i are chosen from a fixed generating set of U.

The corresponding criterion for left modules follows by symmetry and will not be stated separately.

Proof. If U' is pure in U and $a_j = \sum u_i \lambda_{ij}$, let V be the left Λ-module with generators $v_1, \ldots, v_n$ and defining relations $\sum \lambda_{ij} v_j = 0$. Then we have

$$\sum a_j \otimes v_j = \sum u_i \otimes \lambda_{ij} v_j = 0 \qquad \text{in } U \otimes V.$$

From the purity of U' in U we conclude that $\sum a_j \otimes v_j = 0$ in $U' \otimes V$. But this means that $a_j = \sum u_i' \lambda_{ij}$ for some $u_i' \in U'$, by Lemma 2.3.

Conversely, suppose that U' satisfies the condition of the enunciation for all finite sets u_i contained in a fixed generating set of U, and let V be any finitely presented left Λ-module, with generators v_j and relations $\sum \lambda_{ij} v_j = 0 (i = 1, \ldots, m; j = 1, \ldots, n)$. If in $U \otimes V$

$$\sum a_r \otimes b_r = 0,$$

where $a_r \in U'$, $b_r = \sum \alpha_{rj} v_j \in V$, then $\sum a_r \alpha_{rj} \otimes v_j = 0$, hence by Lemma 2.3,

$$\sum a_r \alpha_{rj} = \sum u_s \beta_{si} \lambda_{ij}$$

for some elements $\sum u_s \beta_{si} \in U$. Now $\sum a_r \alpha_{rj} \in U'$; by hypothesis there exist $u_s' \in U'$ such that

$$\sum a_r \alpha_{rj} = \sum u_s' \beta_{si} \lambda_{ij}$$

and hence we have in $U' \otimes V$,

$$\begin{aligned} \sum a_r \otimes b_r = \sum a_r \alpha_{rj} \otimes v_j &= \sum u_s' \beta_{si} \lambda_{ij} \otimes v_j \\ &= \sum u_s' \beta_{si} \otimes \lambda_{ij} v_j \\ &= 0. \end{aligned}$$

This shows that for any finitely presented left Λ-module V, the sequence (2.2) is exact. Now any left Λ-module may be expressed as a direct limit of finitely presented modules, and since direct limits preserve exact sequences, (2.2) is exact for all V and the proof is complete.

From Theorem 2.4 it is not hard to deduce that in the case where Λ is a commutative principal ideal domain, the definition of a pure submodule reduces to the usual one (KAPLANSKY [10], p. 14): a submodule U' is pure in U if and only if $U' \cap U\lambda = U'\lambda$ for all $\lambda \in \Lambda$.

3. Free products and universal products

DEFINITION. A Λ-ring R is said to be the *free product* of the Λ-rings R_ι $(\iota \in I)$ if the R_ι form a family of subrings of R such that

(i) $R_\iota \cap R_\varkappa = \Lambda$ for $\iota \neq \varkappa$,

(ii) if X_ι is a set of generators of R_ι $(\iota \in I)$ then $\bigcup X_\iota$ is a set of generators of R,

(iii) if C_ι is a set of defining relations of R_ι $(\iota \in I)$ (in terms of the set of generators X_ι) then $\bigcup C_\iota$ is a set of defining relations of R.

This corresponds to the definition of a free product of groups with one amalgamated subgroup (cf. B. H. NEUMANN [12]); but since we are using Λ-rings, the ring Λ is embedded canonically in each factor, so that the isomorphisms between the copies of Λ in the different factors are specified automatically.

Given a family of Λ-rings R_ι, there may or may not exist a Λ-ring R which is the free product of the R_ι; in order to discuss conditions for the existence of the free product we shall introduce the "universal product" of the R_ι, which unlike the free product always exists.

THEOREM 3.1. *Let R_ι $(\iota \in I)$ be any family of Λ-rings in the wide sense, then there exists a Λ-ring in the wide sense R, with homomorphisms $\varphi_\iota : R_\iota \to R$ such that*

(i) *R is generated by the $R_\iota \varphi_\iota$,*

(ii) *given any Λ-ring in the wide sense S and any family of homomorphisms $\psi_\iota : R_\iota \to S$, there exists a homomorphism $\psi : R \to S$ satisfying*

$$\psi_\iota = \varphi_\iota \psi .$$

Moreover, R is determined up to isomorphism by (i) *and* (ii).

The Λ-ring in the wide sense R so defined is called the *universal product* of the R_ι, with the *canonical homomorphisms* φ_ι. The construction of R is an instance of BOURBAKI's "problème (U)" (BOURBAKI [4], p. 131; SAMUEL [13]), and we therefore only outline the proof.

We define R as the Λ-ring in the wide sense generated by the elements[8]) of all the rings R_ι, with the defining relations of each R_ι as the defining relations. If the element of R corresponding to an element x of R_ι is denoted by $x\varphi_\iota$, then φ_ι is easily seen to be a homomorphism of R_ι into R, and the φ_ι satisfy (i) by definition. If S is any Λ-ring in the wide sense and $\psi_\iota : R_\iota \to S$ any family of homomorphisms, then the elements

$$x_\iota \varphi_\iota \qquad (x_\iota \in R_\iota,\ \iota \in I) \tag{3.1}$$

form a set of generators of R, the elements

$$x_\iota \psi_\iota \qquad (x_\iota \in R_\iota,\ \iota \in I) \tag{3.2}$$

form a subset of S and by definition every relation holding between the elements (3.1) follows from the defining relations in the different R_ι, and therefore also holds between the elements (3.2). Hence the mapping $\psi : x_\iota \varphi_\iota \to x_\iota \psi_\iota$ of R into S may be extended to a homomorphism of R into S, again denoted by ψ; clearly ψ satisfies (ii). If S, ψ_ι also satisfy (i), (ii), then applying (ii)

[8]) Note that the unit-elements of the different rings all count as the same element here.

with R and S interchanged we obtain a homomorphism $\varphi: S \to R$ such that $\psi_\iota \varphi = \varphi_\iota$. In particular, for every $x_\iota \in R_\iota$, $x_\iota \varphi_\iota \psi \varphi = x_\iota \varphi_\iota$. Since the $x_\iota \varphi_\iota$ ($x_\iota \in R_\iota$, $\iota \in I$) generate R, we have $\psi\varphi = 1$ on a set of generators of R, and hence $\psi\varphi = 1$ on the whole of R. By symmetry $\varphi\psi = 1$ on S, so that $S \cong R$. This establishes the theorem.

Suppose now that the R_ι in Theorem 3.1 are Λ-rings (in the strict sense). Then the free product of the R_ι, when it exists, clearly satisfies (i) and (ii) and is therefore isomorphic to the universal product. Since the universal product is unique up to isomorphism, we have the

COROLLARY. *The free product of any family of Λ-rings, when it exists, is determined up to isomorphism.*

By examining the proof of Theorem 3.1 further, we derive

THEOREM 3.2. *Let R_ι ($\iota \in I$) be a family of Λ-rings and R their universal product. Then the free product of the R_ι exists if and only if*

(i) *the canonical homomorphism $\varphi_\iota: R_\iota \to R$ is a monomorphism,*

(ii) $R_\iota \varphi_\iota \cap R_\varkappa \varphi_\varkappa = \Lambda$ *for* $\iota \neq \varkappa$.

When these conditions hold, R is in fact the free product of the R_ι.

More generally we have the following criterion for the existence of free products. The proof is analogous to that of the (more general) criterion for the existence of generalised free products of groups, by HANNA NEUMANN, and will therefore be omitted (cf. B. H. NEUMANN [12], p. 521).

THEOREM 3.3. *Let R_ι ($\iota \in I$) be a family of Λ-rings. If there exists a Λ-ring S with monomorphisms $\psi_\iota: R_\iota \to S$ such that $R_\iota \psi_\iota \cap R_\varkappa \psi_\varkappa = \Lambda$, for $\iota \neq \varkappa$, then the free product of the R_ι exists.*

We shall find it more convenient to have a criterion in terms of modules; this may be obtained as follows.

Let R be any Λ-ring. By a *right R-module* V we shall understand a unital right Λ-module which is also a right R-module (in the usual sense) in such a way that the right Λ-module structure of V coincides with the right Λ-module structure defined on V by restricting R to Λ. The terms *left R-module*, *R-bimodule* are defined analogously. The module criterion for the existence of free products is now given by

THEOREM 3.4. *Let R_ι ($\iota \in I$) be a family of Λ-rings. If there exists a right Λ-module V such that for each $\iota \in I$, V is a right R_ι-module containing R_ι as submodule*[9]*, and $R_\iota \cap R_\varkappa = \Lambda$ for $\iota \neq \varkappa$, then the free product of the R_ι exists.*

Proof. For each $x \in R_\iota$, let $x\varrho_\iota$ be the endomorphism of V defined by x. If $x \in \Lambda$, then the resulting endomorphism is independent of ι, by definition; if $x \notin \Lambda$, then x belongs to at most one R_ι, by hypothesis, and hence ι is in this case determined by x. Now the right multiplications of V by elements of Λ form a ring isomorphic to Λ, which is a subring of the ring $\mathfrak{E}(V)$ of all endomorphisms of V (qua abelian group). This allows us to regard $\mathfrak{E}(V)$ as a Λ-ring. The mapping $\varrho_\iota: R \to \mathfrak{E}(V)$ is easily seen to be a monomorphism, since $1 \cdot x\varrho_\iota = x$, and so $x\varrho_\iota = 0$ implies $x = 0$. This completes the proof.

[9]) R_ι is regarded as right R_ι-module with respect to right multiplication by R_ι.

Finally we prove a reduction theorem, which shows that when establishing conditions for the existence of free products we may limit ourselves to the case of a finite number of factors[10]).

THEOREM 3.5. *Let R_ι $(\iota \in I)$ be any family of Λ-rings. Then the free product of the R_ι $(\iota \in I)$ exists if and only if the free product of the R_ι $(\iota \in I_0)$ exists for every finite subset I_0 of I.*

Proof. If the free product R of all the R_ι $(\iota \in I)$ exists, then by the definition of the free product, for any subset K of I, the subring of R generated by the $R_\varkappa$ $(\varkappa \in K)$ is the free product of the $R_\varkappa$ $(\varkappa \in K)$.

Conversely, suppose that for every finite subset F of I the free product R_F of the R_ι $(\iota \in F)$ exists. By the first part of the proof, if G is any subset of F, then there is a monomorphism

$$\varrho_G^F : R_G \to R_F,$$

which satisfies the coherence conditions $\varrho_F^F = 1$, $\varrho_H^G \varrho_G^F = \varrho_H^F$ for $H \leqq G \leqq F$. Let R be the direct limit of the system (R_F, ϱ_G^F). Since each ϱ_G^F is a monomorphism, the "projections" $\varrho_F : R_F \to R$ are monomorphisms (cf. EILENBERG & STEENROD [9], p. 221). This defines a canonical embedding of R_ι in R, for each $\iota \in I$. Clearly R satisfies the conditions (i) and (ii) in the definition of the free product. To verify (iii) we note that any relation in R involves only elements from a finite number of rings R_ι and therefore holds in some R_F. Since R_F is a free product, the relation must be a consequence of the relations in the different R_ι $(\iota \in F)$. Thus condition (iii) is also satisfied, and the proof is complete.

4. Sufficient conditions for the existence of the free product

In order to establish the existence of the free product of a given family of Λ-rings R_ι it is enough, by Theorem 3.4, to construct a right Λ-module V which for each $\iota \in I$ is an R_ι-module and contains R_ι as submodule, and such that $R_\iota \cap R_\varkappa = \Lambda$ for $\iota \neq \varkappa$. Moreover, by Theorem 3.5, we may limit ourselves to the case of a finite number of factors. Our first aim (Theorem 4.4) will be to prove the existence, under suitable conditions, of the free product of two factors. The construction is carried out in two steps: First we form a Λ-bimodule U in which R_1 and R_2 are contained as submodules (Lemma 4.1) and then we extend U so as to make it into a right R_i-module $(i = 1, 2)$ (Lemma 4.3). Finally in Lemma 4.2 we obtain a condition which will be used to show that in certain cases the free product is flat.

Throughout this section the Λ-rings occurring in contexts like $R_1 \otimes R_2$ are considered as Λ-bimodules. Further, if R is any Λ-ring, we shall denote the Λ-bimodule R/Λ by $\overline{R}$. If V is any Λ-bimodule, the phrase "V is flat" is ambiguous unless we specify whether V is to be regarded as right or left Λ-module. A similar ambiguity attaches to the expression "the submodule V'

[10]) Cf. DIDIDSE [8], where a corresponding theorem is proved for amalgams of (non-associative) algebras.

is pure in V". We shall remove this ambiguity by saying "V is right flat", "V is left pure in V" etc.; on the rare occasions when there is another coefficient domain S, we distinguish between S-flat modules and Λ-flat modules, but whenever the coefficient domain is not mentioned explicitly, it is understood to be Λ. All this applies in particular when V is a Λ-ring R or the quotient module $\overline{R}$, which is again a Λ-bimodule.

LEMMA 4.1. *Given any Λ-rings R_1 and R_2, put $U = R_1 \otimes R_2$ and define a homomorphism $\vartheta_i : R_i \to U$ $(i = 1, 2)$ by the rule*

$$x\vartheta_1 = x \otimes 1 \qquad (x \in R_1),$$
$$y\vartheta_2 = 1 \otimes y \qquad (y \in R_2).$$

Then ϑ_1, ϑ_2 are monomorphisms and $R_1\vartheta_1 \cap R_2\vartheta_2 = \Lambda$ if and only if the sequence

$$\operatorname{Tor}(R_1, R_2) \to \operatorname{Tor}(\overline{R}_1, \overline{R}_2) \to 0 \tag{4.1}$$

induced by the natural homomorphisms $R_i \to \overline{R}_i$ $(i = 1, 2)$ is exact. In particular, the conclusions hold if $\overline{R}_1$ is right flat, or if $\overline{R}_2$ is left flat.

Proof. By applying the exact homology sequence to the exact sequences

$$0 \to \Lambda \to R_i \to \overline{R}_i \to 0$$

(right modules for $i = 1$, left modules for $i = 2$) and remembering that Λ is free (as right and as left Λ-module) and therefore flat, we obtain the commutative diagram

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & & & \\
 & & \downarrow & & \downarrow & & & & \\
0 & \to & \operatorname{Tor}(R_1, R_2) & \xrightarrow{\alpha} & \operatorname{Tor}(R_1, \overline{R}_2) & \to & R_1 & \xrightarrow{\vartheta_1} & R_1 \otimes R_2 \\
 & & \beta \downarrow & & \delta \downarrow & & \pi \downarrow & & \downarrow \\
0 & \to & \operatorname{Tor}(\overline{R}_1, R_2) & \xrightarrow{\gamma} & \operatorname{Tor}(\overline{R}_1, \overline{R}_2) & \to & \overline{R}_1 & \xrightarrow{\vartheta_1'} & \overline{R}_1 \otimes R_2 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & R_2 & \to & \overline{R}_2 & \to & 0 & & \\
 & & \vartheta_2 \downarrow & & \downarrow & & & & \\
 & & R_1 \otimes R_2 & \to & R_1 \otimes \overline{R}_2 & & & &
\end{array}$$

with exact rows and columns. In terms of this diagram the homomorphism φ in (4.1) is $\alpha\delta = \beta\gamma$, and if this is an epimorphism, the diagram shows that it must be an isomorphism and moreover, $\alpha, \beta, \gamma, \delta$ are all isomorphisms. It follows by exactness that ϑ_1 and ϑ_2 are then monomorphisms. If further,

$$x \otimes 1 = 1 \otimes y \qquad (x \in R_1,\ y \in R_2),$$

then $x\pi \otimes 1 = 0$ and by the exactness of the second row, $x\pi = 0$, i.e. $x \in \Lambda$, therefore $1 \otimes (x - y) = 0$, $x = y \in \Lambda$, which proves that $R_1\vartheta_1 \cap R_2\vartheta_2 = \Lambda$.

Conversely, if ϑ_1, ϑ_2 are monomorphisms then α and β are isomorphisms and so $\alpha\delta = \beta\gamma$ is an isomorphism if and only if ϑ_1' is a monomorphism. Thus if $\beta\gamma$ is not an isomorphism, there exists an $x \in R_1$ such that $x \notin \Lambda$, i.e. $x\pi \neq 0$, but $x\pi \otimes 1 = 0$; in other words we have $x \otimes 1 = \lambda \otimes y = 1 \otimes \lambda y$ $(\lambda \in \Lambda,\ y \in R_2)$. Thus $x \otimes 1 \in R_1\vartheta_1 \cap R_2\vartheta_2$, but $x \notin \Lambda$. This completes the proof.

By Lemma 4.1, if e.g. $\overline{R}_2$ is left flat, we may regard Λ as a submodule of the Λ-bimodule $U = R_1 \otimes R_2$. With this convention we have the following extension of Lemma 4.1.

LEMMA 4.2. *Let R_1 and R_2 be Λ-rings, S a Λ-subring of R_1 and put $U = R_1 \otimes R_2$. If R_1/S is left S-flat and $\overline{R}_2$ is left Λ-flat, then U/R_1 is left S-flat.*

For we may regard R_1 as an (S, Λ)-bimodule and R_2 as a left Λ-module in a natural way. Then U is a left S-module and since $\overline{R}_2$ is left Λ-flat, Λ is left pure in R_2 and we find that

$$U/R_1 \cong R_1 \otimes \overline{R}_2$$

by tensoring the exact sequence

$$0 \to \Lambda \to R_2 \to \overline{R}_2 \to 0$$

on the left with R_1. Thus it remains to show that $R_1 \otimes \overline{R}_2$ is left S-flat, and this follows from Lemma 2.1, since $\overline{R}_2$ is left Λ-flat by hypothesis, and R_1 is left S-flat because both R_1/S and S are. This establishes the lemma.

LEMMA 4.3. *Let U be an (S, Λ)-bimodule, where S is a Λ-ring, and let U' be any submodule of U, with quotient $U'' = U/U'$. Further, let R be a Λ-ring such that the sequence*

$$\operatorname{Tor}(U, R) \to \operatorname{Tor}(U'', \overline{R}) \to 0 \tag{4.2}$$

induced by the natural homomorphisms is exact, and suppose that U' is an (S, R)-bimodule. Then there exists an (S, R)-bimodule

$$W = W(U, U'; R)$$

with an (S, Λ)-bimodule monomorphism $\vartheta: U \to W$ such that the restriction $\vartheta | U'$ is an (S, R)-bimodule monomorphism. If, moreover, the sequence

$$\operatorname{Tor}(U, R) \to \operatorname{Tor}(U, \overline{R}) \to 0 \tag{4.3}$$

is also exact, then

$$W/U \cong U'' \otimes \overline{R}. \tag{4.4}$$

Proof. From the exact sequences

$$0 \to U' \to U \to U'' \to 0 \qquad \text{(right } \Lambda\text{-modules)},$$

$$0 \to \Lambda \to R \to \overline{R} \to 0 \qquad \text{(left } \Lambda\text{-modules)}$$

and the fact that Λ is left flat, we obtain the commutative diagram

$$\begin{array}{ccccccc}
0 \to \operatorname{Tor}(U, R) & \to & \operatorname{Tor}(U, \overline{R}) & & & & \\
\downarrow \alpha & & \downarrow & & & & \\
0 \to \operatorname{Tor}(U'', R) & \xrightarrow{\beta} & \operatorname{Tor}(U'', \overline{R}) & \to & U'' & \to & U'' \otimes R \\
\downarrow & & & & & & \\
U' \otimes R & & & & & & \\
\downarrow & & & & & & \\
U \otimes R & & & & & &
\end{array}$$

with exact rows and columns. By hypothesis $\alpha\beta$ is an epimorphism, hence β is an epimorphism and therefore an isomorphism; it follows that α is also an epimorphism. By exactness we deduce that the sequences

$$0 \to U'' \to U'' \otimes R, \tag{4.5}$$

$$0 \to U' \otimes R \to U \otimes R \tag{4.6}$$

are both exact. Thus we may regard $U' \otimes R$ as a submodule of $U \otimes R$. Because U' is a right R-module, there is a mapping $\varphi: U' \otimes R \to U'$ defined by

$$(u \otimes x)\varphi = ux \qquad (u \in U',\ x \in R),$$

and this is clearly an (S, R)-bimodule homomorphism. Let V be the kernel of φ, so that we have the inclusions $V \subseteq U' \otimes R \subseteq U \otimes R$, qua (S, R)-bimodules. We obtain the commutative diagram[11])

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & V & \to & U' \otimes R & \to & U' & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & V & \to & U \otimes R & \xrightarrow{\alpha} & W & \to & 0 \\
 & & \downarrow & & \downarrow \beta & & \downarrow & & \\
 & & 0 & \to & U'' \otimes R & \to & U'' \otimes R & \to & 0 \\
 & & & & \downarrow & & \downarrow & & \\
 & & & & 0 & & 0 & &
\end{array}$$

with exact rows and columns. Here $W = (U \otimes R)/V$ is an (S, R)-bimodule since both $U \otimes R$ and V are (S, R)-bimodules. Combining the homomorphism $u \to u \otimes 1$ of U into $U \otimes R$ with the homomorphism $\alpha: U \otimes R \to W$ given by the diagram, we obtain a homomorphism $\vartheta: U \to W$, whose restriction to U' is an (S, R)-bimodule monomorphism, again from the diagram. To show that ϑ itself is a monomorphism, let $u\vartheta = 0$ $(u \in U)$, then $(u \otimes 1)\alpha = 0$, hence $u \otimes 1$ is the image of an element from V and the diagram shows that $(u \otimes 1)\beta = 0$. If $u \to u^*$ is the natural homomorphism $U \to U''$, then this may be written $u^* \otimes 1 = 0$, whence by (4.5), $u^* = 0$, i.e. $u \in U'$. Since $\vartheta \mid U'$ has already been shown to be a monomorphism, the result follows.

Now suppose that (4.3) holds; then the sequence

$$0 \to U \to U \otimes R \tag{4.7}$$

is exact, and combining this with (4.6), we obtain the commutative diagram

$$\begin{array}{ccccccccc}
 & & & & 0 & & & & \\
 & & & & \downarrow & & & & \\
 & & U' & \to & U' \otimes R & \to & U' \otimes \overline{R} & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & U & \to & U \otimes R & \to & U \otimes \overline{R} & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & U'' & \to & U'' \otimes R & \to & U'' \otimes \overline{R} & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

[11]) CARTAN & EILENBERG [5], p. 16, ex. 1.

with exact rows and columns. From it we obtain the exact sequence

$$0 \to U + U' \otimes R \to U \otimes R \to U'' \otimes \overline{R} \to 0.$$

By (4.7) we may regard U as embedded in $U \otimes R$; then the fact that ϑ is a monomorphism means that $U \cap V = 0$, while $U + V = U + U' \otimes R$, as is easily verified. It follows that

$$U'' \otimes \overline{R} \cong (U \otimes R)/(U + V) \cong (U \otimes R)/V/(U + V)/V \cong W/U,$$

and this establishes the lemma.

We note that if in Lemma 4.3, $\overline{R}$ is left flat, then both the sequences (4.2) and (4.3) are exact; if moreover U'' is left S-flat then W/U is left S-flat. This is seen by applying Lemma 2.1 to the result (4.4).

A set of sufficient conditions for the existence of the free product of two factors can now be given.

THEOREM 4.4. *Let R_1 and R_2 be any Λ-rings and let S be any Λ-subring of $\overline{R}_1$. If* (i) *$\overline{R}_i = R_i/\Lambda$ is left flat ($i = 1, 2$) and* (ii) *R_1/S is left S-flat, then the free product R of R_1 and R_2 exists, with S as Λ-subring, such that R/S is again left S-flat.*

Proof. First we form $U = R_1 \otimes R_2$; by Lemma 4.1, each R_i is embedded in a natural way in U and

$$R_1 \cap R_2 = \Lambda \quad \text{in } U. \tag{4.8}$$

Put $W_1 = R_1$ and $W_2 = U$, then W_2 is an (S, Λ)-bimodule and W_1 a submodule of W_2 which is also a right R_1-module. Now let $n > 2$ and suppose that an ascending chain of (S, Λ)-bimodules

$$W_1 \subseteq W_2 \subseteq \cdots \subseteq W_{n-1}$$

has already been defined, where W_r is an R_1-module if r is odd and an R_2-module if r is even. Then we put

$$W_n = W(W_{n-1}, W_{n-2}; R_i) \qquad (i \equiv n \pmod 2),$$

where the module on the right is defined as in Lemma 4.3. This lemma also shows that there is a natural embedding of W_{n-1} in W_n (because $\overline{R}_i$ is left flat), that W_n is an R_i-module with W_{n-2} as submodule, and that

$$W_n/W_{n-1} \cong (W_{n-1}/W_{n-2}) \otimes \overline{R}_i. \tag{4.9}$$

Let V be the union of the ascending sequence of modules W_n, then V is an (S, Λ)-bimodule and qua Λ-module, V contains R_1 and R_2 as submodules such that $R_1 \cap R_2 = \Lambda$, by (4.8). If we can show that V is an R_i-module with R_i as submodule, for $i = 1, 2$, then we may apply Theorem 3.4 and conclude that the free product of R_1 and R_2 exists. Let $v \in V$, $x \in R_i$, then $v \in W_n$ for some n, and hence $v \in W_\nu$ for all $\nu > n$. If we choose ν to be of the form $2k + i$, for some k, then W_ν is an R_i-module with R_i as submodule, and therefore vx is defined in W_ν. The definition is independent of the choice of k, since W_μ is a submodule of W_ν, qua R_i-module, whenever $\nu - \mu = 2k > 0$. The product

vx is therefore uniquely defined in V and it is easily verified that this definition makes V an R_1-module. This establishes the existence of the free product, and hence the first part of the theorem.

For the remaining part we note that $W_2/W_1 = U/R_1$ is left S-flat, by Lemma 4.2. Using (4.9) we find, by induction on n, that W_n/W_{n-1} is left S-flat for all n, and hence $W_n/W_1 = W_n/R_1$ is left S-flat; since R_1/S is S-flat by hypothesis, W_n/S is S-flat, and going to the limit we find that V/S is S-flat. Now V is the free right R-module on one free generator, namely the unit-element of Λ. It follows from the equation

$$s1 \cdot x = 1 \cdot sx \qquad (s \in S,\ x \in R)$$

that V and R are isomorphic, qua left S-modules, and moreover S itself, which is a submodule of both, corresponds elementwise to itself in the isomorphism. Hence $V/S \cong R/S$ and this proves that R/S is left S-flat. This completes the proof of the theorem.

If in the theorem we take $S = \Lambda$ or $S = R_1$, the hypothesis that R_1/S be left S-flat is certainly verified, and we obtain the following corollaries:

COROLLARY 1. *If R_1 and R_2 are any Λ-rings such that $\overline{R}_i$ is left flat $(i = 1, 2)$, then the free product R of R_1 and R_2 exists and $\overline{R}$ is left flat.*

COROLLARY 2. *If R_1, R_2 and R are as in Cor. 1, then R/R_1 is left R_1-flat.*

From Cor. 1 we can easily deduce the existence of the free product of any family of Λ-rings R_ι such that each $\overline{R}_\iota$ is left (or right) flat. First let $R_1, \ldots, R_n$ be any finite set of Λ-rings such that each $\overline{R}_i$ is left flat, and suppose that it has already been shown that the free product T of $R_1, \ldots, R_{n-1}$ exists and that $\overline{T}$ is left flat. By Theorem 4.4 Cor. 1, the free product R of T and R_n exists and $\overline{R}$ is left flat. Now it is easily verified, using Theorem 3.1 Cor., that R is the free product of $R_1, \ldots, R_n$. It follows by induction that the free product of any finite family of rings R_ι satisfying the condition that $\overline{R}_\iota$ be left flat, exists and satisfies the same condition. We may now apply Theorem 3.5 to prove the existence of the free product of any family of rings satisfying the above condition. Since flatness is preserved under direct limits, $\overline{R}$ is again left flat. Of course the corresponding assertion with "left" replaced by "right" also holds, by symmetry. Thus we have proved

THEOREM 4.5. *Let R_ι $(\iota \in I)$ be any family of Λ-rings. If for each $\iota \in I$, the module R_ι/Λ is left flat, then the free product R of the R_ι exists and R/Λ is again left flat. A similar assertion holds with "left" replaced by "right" throughout.*

The further assertion of Theorem 4.4, concerning S, could also be extended to arbitrary families of rings, but this may be left to the reader. We note one further consequence of Theorem 4.4, giving some information about the structure of the free product.

THEOREM 4.6. *If R_1 and R_2 are any Λ-rings and R_i/Λ is left flat $(i = 1, 2)$ so that the free product R of R_1 and R_2 exists, then R has a filtration*

$$A_1 \subseteq A_2 \subseteq \cdots, \quad \cup A_n = R \qquad (A_i A_j \subseteq A_{i+j})$$

where the A_i are Λ-bimodules and

$$A_n/A_{n-1} \cong R_1 \otimes \overline{R}_2 \otimes \overline{R}_1 \otimes \cdots \otimes \overline{R}_i \qquad (n \text{ factors}, \ i \equiv n \pmod 2).)$$

This follows by taking A_n to be the submodule of R corresponding to W_n in the isomorphism of V and R established in the proof of Theorem 4.4, and applying (4.9).

It is known that any Λ-module is flat provided that Λ is a regular ring, i.e. if for any $\lambda \in \Lambda$, the equation

$$\lambda \xi \lambda = \lambda$$

has a solution in Λ (AUSLANDER [1]). We therefore obtain the following corollary to Theorem 4.5.

THEOREM 4.7. *Let Λ be a regular ring; then for any family of Λ-rings the free product exists.*

In particular the free product of Λ-rings always exists if Λ is a semisimple ring. By further specialising Λ we obtain the

COROLLARY. *The free product of any family of Λ-rings exists if Λ is a division ring.*

5. Necessary conditions for the existence of free products

The problem of obtaining conditions for the existence of the free product can be treated on three levels. We may study (i) conditions on the Λ-rings R_ι ($\iota \in I$) for their free product to exist, (ii) conditions on the Λ-ring R for the free product of R with any Λ-ring S to exist, and (iii) conditions on Λ for the free product of any family of Λ-rings to exist. Whereas the sufficient conditions given in the last section were of the types (iii) and (i), we shall confine ourselves to the cases (ii) and (iii) in establishing necessary conditions[12]. We begin with a necessary condition of type (ii).

THEOREM 5.1. *Let R be any Λ-ring. If the free product of R with any Λ-ring exists, then for any right Λ-module U and any left Λ-module V, the homomorphism $\varphi: U \otimes V \to U \otimes R \otimes V$ defined by*

$$(u \otimes v)\varphi = u \otimes 1 \otimes v \tag{5.1}$$

is a monomorphism.

Proof. Suppose that the condition is not satisfied and let U, V be a right and left Λ-module respectively with elements $u_i \in U$ and $v_i \in V$ such that

$$\sum u_i \otimes v_i \neq 0 \quad \text{in } U \otimes V, \quad \text{but} \quad (\sum u_i \otimes v_i)\varphi = 0. \tag{5.2}$$

Put ${}_\Lambda U = \Lambda \otimes_Z U$, $V_\Lambda = V \otimes_Z \Lambda$, where Z is the ring of integers, and let $W = {}_\Lambda U \oplus V_\Lambda$, so that W is a Λ-bimodule. Now let S be the tensor algebra[13] over Λ on the Λ-bimodule W. The terms of degree two in S form a module $W \otimes W$ which contains ${}_\Lambda U \otimes V_\Lambda$ as a direct summand. Since S is graded, it

[12] Cf. however Lemma 4.1, which contains a necessary condition of type (i).

[13] Cf. e.g. CHEVALLEY [6]; the discussion there is confined to Λ-algebras, but may be extended to Λ-rings without difficulty.

contains $W \otimes W$ as a direct summand, and hence ${}_\Lambda U \otimes V_\Lambda$ is a direct summand in S. It follows that

$$\sum u_i \otimes v_i \neq 0 \quad \text{in } S. \tag{5.3}$$

Now suppose that the free product P of R and S exists. Then there is a trilinear mapping of $S \times R \times S$ into P, defined by $(x, y, z) \to xyz$ ($y \in R$, $x, z \in S$). Hence there is a module homomorphism of $S \otimes R \otimes S$ into P, defined by $x \otimes y \otimes z \to xyz$. By (5.2), $\sum u_i \otimes 1 \otimes v_i = 0$ and so $\sum u_i v_i = 0$ in P; a comparison with (5.3) shows that the natural homomorphism $S \to P$ is not an embedding. Therefore the free product of R and S does not exist, and the theorem follows.

The condition of Theorem 5.1 may be reformulated as follows: Suppose that U and V are finitely presented modules, with generating sets u_r, v_s and defining relations $\sum u_r \mu_{rj} = 0$, $\sum \lambda_{is} v_s = 0$ respectively. Let $w = \sum u_r \xi_{rs} \otimes v_s$ be an element of the kernel of φ, given by (5.1), then by Lemma 2.3, since the u_r generate U, we have

$$\sum u_r \xi_{rs} \otimes 1 = \sum u_r \otimes a_{ri} \lambda_{is} \qquad \text{in } U \otimes R,$$

for some $a_{ri} \in R$. Applying Lemma 2.3 again, this time to U, we find that

$$\xi_{rs} = \sum a_{ri} \lambda_{is} + \sum \mu_{rj} b_{js}, \tag{5.4}$$

for some $b_{js} \in R$. Conversely, if (5.4) holds, then $(\sum u_r \xi_{rs} \otimes v_s)\varphi = 0$, as is easily seen. On the other hand, the equation

$$\sum u_r \xi_{rs} \otimes v_s = 0$$

holds in $U \otimes V$ if and only if there exist $\alpha_{ri}, \beta_{js} \in \Lambda$ such that

$$\xi_{rs} = \sum \alpha_{ri} \lambda_{is} + \sum \mu_{rj} \beta_{js} \tag{5.5}$$

as follows by taking $R = \Lambda$ in the preceding argument. Therefore if φ is to be a monomorphism, any equation of the form (5.4) must imply the corresponding equation (5.5) in Λ. This proves

THEOREM 5.2. *Let R be a Λ-ring such that the free product of R with any Λ-ring exists. Then for any finite sets of elements $\xi_{rs}, \lambda_{is}, \mu_{rj} \in \Lambda$ and $a_{ri}, b_{js} \in R$, the equation*

$$\xi_{rs} = \sum a_{ri} \lambda_{is} + \sum \mu_{rj} b_{js}$$

implies the existence of $\alpha_{ri}, \beta_{js} \in \Lambda$ such that

$$\xi_{rs} = \sum \alpha_{ri} \lambda_{is} + \sum \mu_{rj} \beta_{js}.$$

COROLLARY. *Given a Λ-ring R, if the free product of R with any Λ-ring exists, then Λ is right pure and left pure in R.*

The corollary follows by taking in turn $\lambda_{is} = 0$ and $\mu_{rj} = 0$ in the theorem and applying Theorem 2.4. Of course it also follows directly from Theorem 5.1.

From the corollary we readily obtain a necessary condition on Λ. We could of course use the possibly more general condition of Theorem 5.2, but

this is rather more difficult to handle, and we shall have no more to say about it.

THEOREM 5.3. *Given any ring Λ with 1, if any free product of Λ-rings exists, then Λ must be right pure and left pure in any Λ-ring.*

For brevity we shall say that Λ is *right pure*, if Λ is right pure in every Λ-ring; the expression "left pure" is defined similarly. When Λ is an algebra over a field, it will follow from Theorem 5.5 that Λ is right (left) pure if and only if it is right (left) pure in every Λ-bimodule containing Λ as a submodule, but we shall not make use of this result.

Our next object is to give a criterion for a ring to be left pure which will enable us to construct rings which are not left pure. Of course the same considerations apply to the right purity of a ring. We shall confine our attention to the case where Λ is an algebra over a field F. It is clear that any Λ-ring may then also be considered as an algebra over F. The argument could be carried through in the case where Λ has a right operator domain with respect to which it is completely reducible, but we shall not enter into this refinement. We begin with a theorem which is of some interest in itself.

THEOREM 5.4. *Let Λ be an algebra with 1 over a field F. Given any elements $\lambda_{ij}, \alpha_i \in \Lambda$ $(i = 1, \ldots, m;\ j = 1, \ldots, n)$, there exists a Λ-ring R with elements t_j $(j = 1, \ldots, n)$ such that*

$$\sum \lambda_{ij} t_j = \alpha_i \qquad (i = 1, \ldots, m), \tag{5.6}$$

if and only if, for all sets of elements $\xi_i \in \Lambda$,

$$\sum \xi_i \lambda_{ij} = 0 \quad (j = 1, \ldots, n) \quad \textit{implies} \quad \sum \xi_i \alpha_i = 0. \tag{5.7}$$

Proof. It is clear that condition (5.7) is necessary. In order to prove the sufficiency we require a lemma.

LEMMA. *If (5.7) holds for all $\xi_i \in \Lambda$, then there exists a Λ-ring R with an element t such that for all $\xi_i \in \Lambda$,*

$$\sum \xi_i \lambda_{ij} = 0 \quad (j = 2, \ldots, n) \quad \textit{implies} \quad \sum \xi_i (\alpha_i - \lambda_{i1} t) = 0.$$

To prove the lemma, we put

$$I = \{\textstyle\sum \xi_i \lambda_{i1} \mid \xi_i \in \Lambda \text{ such that } \sum \xi_i \lambda_{ij} = 0 \text{ for } j = 2, \ldots, n\}.$$

Thet set I is easily seen to be a left ideal of Λ. Let I' be a complementary subspace of I in Λ, and define a linear transformation t on Λ as follows: given any $\mu \in \Lambda$, let

$$\mu = \sum \xi_i \lambda_{i1} + \mu' \qquad (\mu' \in I',\ \textstyle\sum \xi_i \lambda_{i1} \in I). \tag{5.8}$$

By definition of I' such a decomposition is possible, and μ' is uniquely determined by μ. The elements ξ_i are not uniquely determined by μ, but if we also have

$$\mu = \sum \xi_i' \lambda_{i1} + \mu'$$

then $\sum(\xi_i - \xi_i')\lambda_{i1} = 0$, and since $\sum(\xi_i - \xi_i')\lambda_{ij} = 0$ for $j = 2, \ldots, n$, by definition of I, it follows from (5.7) that

$$\sum \xi_i \alpha_i - \sum \xi_i' \alpha_i = 0.$$

Hence if we put

$$\mu t = \sum \xi_i \alpha_i + \mu', \tag{5.9}$$

the element μt is uniquely determined by μ. Since I and I' are subspaces it follows that $\mu \to \mu t$ is a linear transformation.

Now consider the ring $\mathfrak{L} = \mathfrak{L}(\Lambda)$ of all linear transformations of Λ qua space over F. The right multiplications by elements of Λ form a subring of $\mathfrak{L}$ isomorphic with Λ, and identifying the right multiplications with the elements of Λ, we may regard Λ as a subring of $\mathfrak{L}$ and hence $\mathfrak{L}$ as a Λ-ring. If $\mu = \sum \xi_i \lambda_{i1}$ is any element of I, then

$$\mu t = (\sum \xi_i \lambda_{i1}) t = \sum \xi_i \alpha_i,$$

by (5.9), hence $\sum \xi_i (\alpha_i - \lambda_{i1} t) = 0$ in $\mathfrak{L}$, and this establishes the lemma.

We now complete the proof of Theorem 5.4 by induction on n. For $n = 0$, the hypothesis states that $\sum \xi_i \alpha_i = 0$ for all $\xi_i \in \Lambda$. Taking $\xi_i = \delta_{ik}$, the Kronecker delta, we find that $\alpha_i = 0$ $(i = 1, \ldots, m)$ and hence (5.6) holds with $t_j = 0$. Now let $n > 0$; by the lemma there is a Λ-ring R_1 containing an element t_1 such that

$$\sum \xi_i \lambda_{ij} = 0 \quad (j = 2, \ldots, n) \quad \text{implies} \quad \sum \xi_i (\alpha_i - \lambda_{i1} t_1) = 0.$$

Moreover, R_1 is again an algebra over F. Applying the induction hypothesis to R_1 we find a ring R containing R_1 and elements $t_2, \ldots, t_n$ such that

$$\alpha_i - \lambda_{i1} t_1 = \sum_{j=2}^{n} \lambda_{ij} t_j \quad (i = 1, \ldots, m),$$

in other words,

$$\alpha_i = \sum_{j=1}^{n} \lambda_{ij} t_j.$$

This completes the proof of the theorem.

As an illustration of Theorem 5.4, we note that if R is any algebra with 1 and $a \in R$ is not a right zero-divisor, then a right inverse of a may be adjoined to R.

We use Theorem 5.4 to establish the following criterion for an algebra to be left pure.

THEOREM 5.5. *An algebra Λ with 1 is left pure if and only if, given any $\lambda_{ij}, \alpha_i \in \Lambda$ $(i = 1, \ldots, m;\ j = 1, \ldots, n)$ such that for all $\xi_i \in \Lambda$,*

$$\sum \xi_i \lambda_{ij} = 0 \quad \textit{implies} \quad \sum \xi_i \alpha_i = 0, \tag{5.10}$$

there exist $\beta_j \in \Lambda$ such that

$$\alpha_i = \sum \lambda_{ij} \beta_j. \tag{5.11}$$

Proof. Suppose that the condition is satisfied. Given any Λ-ring R, let

$$\alpha_i = \sum \lambda_{ij} t_j \qquad (\alpha_i, \lambda_{ij} \in \Lambda,\ t_j \in R). \tag{5.12}$$

Then it clear that (5.10) holds for all $\xi_i \in \Lambda$, and so (5.11) holds for some $\beta_j \in \Lambda$. But this just means that Λ is left pure in R, by Theorem 2.4. Conversely, suppose that Λ is left pure in every Λ-ring R, and let (5.10) be satisfied. Then by Theorem 5.4, there is a Λ-ring R with elements t_j satisfying (5.12), and since Λ is left pure in R, (5.11) follows. Therefore the condition of the theorem holds and the proof is complete.

The remark following Theorem 5.4, together with Theorem 5.5 shows that in a left pure algebra Λ any element which is not a right zero-divisor must have a right inverse. Thus e.g. the ring $F[x]$ of polynomials in a variable x over a field F is not left pure. Now Theorem 5.3 shows that there are $F[x]$-rings whose free product does not exist.

6. Rings without a unit-element

Up to this point we have dealt only with the class of rings with a unit-element. By regarding the unit-element as a constant (i.e. 0-ary) operator we obtained a product in which the unit-elements of all rings coincide. We shall now indicate briefly the modifications necessary in the general case. It should be emphasised that some or all of the rings occurring may still contain a unit-element, but these unit-elements will not be identified in the free product, unless they belong to the common subring Λ.

Given any ring R, we shall denote by R^1 the ring obtained by adjoining a unit-element 1 to R. This construction may be carried out whether or not R has a unit-element; it is to be understood that any unit-element already existing in R is denoted by a symbol other than 1. If the unit-element of R is e say, then $e1 = e \neq 1$, so that e is not unit-element in R^1. In order to recover R from R^1 we make the following

DEFINITION. A *1-ring* is a ring T with unit-element 1 and an ideal R such that $T/R \cong Z$, where Z is the ring of integers. R is called the *ideal associated with* T.

Clearly, for any ring R, R^1 is a 1-ring with associated ideal R, and if T is any 1-ring with associated ideal R, then $T \cong R^1$ (cf. [7]).

If Λ is any ring and R is any ring containing Λ as a subring, then it is clear that R^1 is a Λ^1-ring. Therefore if we have a family R_ι ($\iota \in I$) of rings each containing Λ as a subring, then we can apply the theorems of §§ 3–5 to the rings R_ι^1. All that is needed is a proof that we can drop the 1 from the universal product. This follows from

THEOREM 6.1. *If T_ι ($\iota \in I$) is any family of Λ^1-rings, and each T_ι is a 1-ring, then the universal product of the T_ι is a 1-ring.*

Proof. Let R_ι be the ideal associated with T_ι, and let T be the universal product of the T_ι. We define the ring Z of integers as a Λ^1-ring in the wide sense by putting

$$(m + \lambda) n = n(m + \lambda) = mn \qquad (m, n \in Z,\ \lambda \in \Lambda).$$

Then the natural homomorphism $\psi_i: T_i \to Z$ is a Λ^1-homomorphism and hence there is a homomorphism $\psi: T \to Z$ such that $\psi_i = \varphi_i \psi$, where φ_i is the canonical homomorphism $T_i \to T$. Since each ψ_i is an epimorphism, so is ψ, and the theorem follows.

It is clear from the proof that the ideal R associated with the universal product T contains the images $R_i\varphi_i$ of the ideals associated with the T_i. We may therefore regard R as the universal product of the R_i; moreover, the free product of the R_i (with Λ as common subring) exists if and only if the free product of the R_i^1 qua Λ^1-rings exists.

Bibliography

[1] AUSLANDER, M.: On regular group rings. Proc. Amer. Math. Soc. **8**, 658–664 (1957). — [2] BIRKHOFF, G.: Lattice theory. New York 1948. — [3] BOURBAKI, N.: Algèbre, chap. II. Paris 1947. — [4] BOURBAKI, N.: Algèbre, chap. III, appendice III. Paris 1948. — [5] CARTAN, H., and S. EILENBERG: Homological algebra. Princeton 1956. — [6] CHEVALLEY, C.: The construction and study of certain important algebras. Tokyo 1955. — [7] COHN, P. M.: On a class of simple rings. Mathematika **5**, 103–117 (1958). — [8] DIDIDSE, Z. E.: Non-associative free sums of algebras with amalgamated subalgebras (Russian). Mat. Sbornik **43** (85), 379–396 (1957). — [9] EILENBERG, S., and N. STEENROD: Foundations of algebraic topology. Princeton 1952. — [10] KAPLANSKY, I.: Infinite abelian groups. Ann Arbor 1954. — [11] KUROŠ, A. G.: Non-associative free algebras and free products of algebras. Rec. Math. (Mat. Sbornik) **20** (62), 239–262 (1947). — [12] NEUMANN, B. H.: An essay on free products of groups with amalgamations. Phil. Trans. Roy. Soc., Ser. A **246**, 503–554 (1954). — [13] SAMUEL, P.: On universal mappings and free topological groups. Bull. Amer. Math. Soc. **54**, 591–598 (1948). — [14] SCHREIER, O.: Die Untergruppen der freien Gruppen. Abh. math. Sem. Hamburg **5**, 161–183 (1927). — [15] SIKORSKI, R.: Products of abstract algebras. Fund. Math. **39**, 211–228 (1953).

The University, Manchester, England

(Eingegangen am 26. Januar 1959)

KOPPELMAN, W., and J. D. PINCUS
Math. Zeitschr. 71, 399–407 (1959)

Spectral Representations for Finite Hilbert Transformations

By

W. KOPPELMAN and J. D. PINCUS*)

1. INTRODUCTION. It is well-known that the transformation

$$H x(\lambda) \equiv \frac{1}{\pi i} \int_{-\infty}^{\infty} \frac{x(\mu)}{\mu - \lambda} d\mu \tag{1.1}$$

acting on functions $x(\lambda)$ of class $L^2(-\infty, \infty)$ is a unitary symmetric operator on the Hilbert space $L^2(-\infty, \infty)$ [9]. (The integral here is to be interpreted as a Cauchy principal value at $\mu = \lambda$.) We shall refer to the operator H as the Hilbert transformation. The spectrum of the operator H consists just of the points ± 1. Under the application of the Fourier-Plancherel transformation

$$y(\xi) = \frac{1}{\sqrt{2\pi}} \operatorname*{l.i.m.}_{N \to \infty} \int_{-N}^{N} e^{i\xi\lambda} x(\lambda)\, d\lambda \equiv U x(\lambda) \tag{1.2}$$

one obtains a particularly simple representation of the operator H. It can be shown that if $y(\xi) = U x(\lambda)$, then $-\operatorname{sgn} \xi \cdot y(\xi) = U H x(\lambda)$ [9].

In this paper we shall study operators of the form

$$T_{(a,b)} x(\lambda) \equiv \frac{1}{\pi i} \int_{a}^{b} \frac{x(\mu)}{\mu - \lambda} d\mu, \quad a < b, \quad a < \lambda < b, \tag{1.3}$$

acting on functions $x(\lambda)$ of class $L^2(a, b)$, where at least one of the limits of integration a, b is finite. The operators $T_{(a,b)}$ will be called finite Hilbert transformations if both a and b are finite and semi-infinite Hilbert transformations if one of the two limits of integration is infinite. Operators of the form $T_{(a,b)}$ appear in integral equations of airfoil theory [7], which have been studied by various authors [1, 3, 4, 6, 8, 10, 11].

Because of the norm-preserving property of the Hilbert transformation, one can readily deduce the following lemma:

LEMMA 1.1. *The operators $T_{(a,b)}$ are bounded on $L^2(a, b)$ with norm $\|T_{(a,b)}\| \leqq 1$.*

*) This paper was prepared under contract no. DA-30-069-ORD-1258 with the Office of Ordnance Research, United States Army, at the Institute of Mathematical Sciences of New York University.

We shall show that the Fourier-Plancherel theory also leads to a simple representation for the finite and semi-infinite Hilbert transformations. From this representation one immediately concludes that the operators $T_{(a,b)}$ possess a simple continuous spectrum from -1 to 1 and hence that $\|T_{(a,b)}\| = 1$.

2. A GENERALIZED EIGENFUNCTION FOR THE OPERATOR $T_{(a,b)}$. In order to obtain eigenfunctions for the operator $T_{(a,b)}$, we must study the resolvent equation

$$\xi\, x(\lambda) - \frac{1}{\pi i}\int_a^b \frac{x(\mu)}{\mu-\lambda}\, d\mu = 0. \tag{2.1}$$

In the present discussion we restrict ourselves to the case where both a and b are finite. We now use the following well-known method, devised by CARLEMAN [2], to reduce the singular integral equation (2.1) to a boundary value problem of function theory:

Set

$$X(z) = \frac{1}{2\pi i}\int_a^b \frac{x(\mu)}{\mu - z}\, d\mu, \tag{2.2}$$

where $x(\mu)$ is a solution of (2.1). If $x(\mu)$ satisfies a Hölder condition on every proper subinterval of the interval (a, b), then $X(z)$ satisfies the Plemelj relations

$$\begin{cases} X(\lambda + i0) - X(\lambda - i0) = x(\lambda), \\ X(\lambda + i0) + X(\lambda - i0) = \dfrac{1}{\pi i}\displaystyle\int_a^b \frac{x(\mu)}{\mu-\lambda}\, d\mu, \quad a < \lambda < b \end{cases} \tag{2.3}$$

and the condition

$$X(\infty) = 0. \tag{2.4}$$

If, furthermore, $x(\lambda)$ is of the form $\frac{x^*(\lambda)}{(\lambda-a)^\gamma}$, $\frac{x^{**}(\lambda)}{(\lambda-b)^\gamma}$, $0 \leqq \mathrm{Re}\,\gamma < 1$, near the ends a and b respectively, where $x^*(\lambda)$ satisfies a Hölder condition near and at a and $x^{**}(\lambda)$ satisfies a Hölder condition near and at b, then

$$X(z) = O\big((z-a)^{-\alpha}\big), \quad \alpha < 1, \quad \text{near } a \tag{2.5}$$

and

$$X(z) = O\big((z-b)^{-\beta}\big), \quad \beta < 1, \quad \text{near } b. \text{ [6]} \tag{2.6}$$

From the Plemelj relations (2.3) and equation (2.1), we see that $X(z)$ must satisfy the barrier equation

$$(\xi - 1)\, X(\lambda + i0) - (\xi + 1)\, X(\lambda - i0) = 0, \quad a < \lambda < b. \tag{2.7}$$

Conversely, suppose that $X(z)$ is a solution of the barrier equation (2.7), such that $X(\lambda \pm i0)$ is Hölder continuous on every proper subinterval of (a, b). Assume, furthermore, that $X(z)$ also satisfies conditions (2.4), (2.5), and (2.6). Form the function $x(\lambda) = X(\lambda + i0) - X(\lambda - i0)$. Then $x(\lambda)$ also satisfies such

a Hölder condition and the end-point conditions listed above. Thus the function $Y(z)=\frac{1}{2\pi i}\int_a^b \frac{x(\mu)}{\mu-z}\,d\mu$ satisfies (2.3), (2.4), (2.5) and (2.6). Consider the function $X(z)-Y(z)$. We have

$$\{X(\lambda+i0)-Y(\lambda+i0)\}-\{X(\lambda-i0)-Y(\lambda-i0)\}=0$$

or $\lambda \neq a, b$. Hence $X(z)-Y(z)$ is analytic for $z \neq a, b$. Because of the conditions (2.5), (2.6), $X(z)-Y(z)$ has removable singularities at $z=a$, $z=b$. From (2.4) we have $X(\infty)-Y(\infty)=0$. Thus $X(z)-Y(z)$ is regular in the whole plane and vanishes at ∞. By LIOUVILLE's theorem $X(z)-Y(z)\equiv 0$. Thus the solution $X(z)$ of the barrier equation (2.7) has the form $\frac{1}{2\pi i}\int_a^b \frac{x(\mu)}{\mu-z}\,d\mu$. Applying the Plemelj relations, we see that $x(\lambda)$ is a solution of (2.1).

If ξ is real and $|\xi|<1$, the functions

$$E_+(\xi,z)=e^{\frac{1}{2\pi i}\left\{\log\frac{\xi+1}{1-\xi}+\pi i\right\}\int_a^b \frac{d\mu}{\mu-z}} \tag{2.8}$$

$$E_-(\xi,z)=e^{\frac{1}{2\pi i}\left\{\log\frac{\xi+1}{1-\xi}-\pi i\right\}\int_a^b \frac{d\mu}{\mu-z}} \tag{2.9}$$

both are solutions of the barrier equation (2.7), which satisfy the conditions (2.5) and (2.6). Furthermore $E_+(\xi,\infty)=E_-(\xi,\infty)=1$. Thus $E_+(\xi,z)-E_-(\xi,z)$ is a solution of the barrier equation (2.7), which satisfies *all* of our conditions. According to our procedure above, the function

$$\{E_+(\xi,\lambda+i0)-E_-(\xi,\lambda+i0)\}-\{E_+(\xi,\lambda-i0)-E_-(\xi,\lambda-i0)\} \tag{2.10}$$

is a solution of the equation (2.1). Let us evaluate this function explicitly:

$$E_\pm(\xi,\lambda+i0)=e^{\left\{\log\frac{\xi+1}{1-\xi}\pm\pi i\right\}\left\{\frac{1}{2}+\frac{1}{2\pi i}\log\frac{b-\lambda}{\lambda-a}\right\}}$$

$$E_\pm(\xi,\lambda-i0)=e^{\left\{\log\frac{\xi+1}{1-\xi}\pm\pi i\right\}\left\{-\frac{1}{2}+\frac{1}{2\pi i}\log\frac{b-\lambda}{\lambda-a}\right\}}.$$

The function (2.10) is thus equal to

$$i\,e^{\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\lambda-a}{b-\lambda}\right)}\left\{\sqrt{\frac{\xi+1}{1-\xi}}\left(\sqrt{\frac{b-\lambda}{\lambda-a}}+\sqrt{\frac{\lambda-a}{b-\lambda}}\right)+\right.$$

$$\left.+\sqrt{\frac{1-\xi}{1+\xi}}\left(\sqrt{\frac{b-\lambda}{\lambda-a}}+\sqrt{\frac{\lambda-a}{b-\lambda}}\right)\right\}=2i(b-a)\,\frac{e^{\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\lambda-a}{b-\lambda}\right)}}{\sqrt{(1-\xi^2)(\lambda-a)(b-\lambda)}}.$$

Thus we have

LEMMA 2.1. *If ξ is real with $|\xi|<1$, then the function* $\frac{e^{\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\lambda-a}{b-\lambda}\right)}}{\sqrt{(\lambda-a)(b-\lambda)}}$ *is a solution of equation* (2.1).

3. THE SPECTRAL REPRESENTATION FOR THE FINITE HILBERT TRANSFORMATION. Using the Fourier-Plancherel theory and an appropriate change of variables, we can now prove

THEOREM 3.1. *The formulae*

$$(3.1)\quad f(\lambda)=\frac{1}{\pi}\sqrt{\frac{b-a}{2}}\operatorname*{l.i.m.}_{\delta\to 0,\,\delta>0}\int_{-1+\delta}^{1-\delta}\frac{e^{\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\lambda-a}{b-\lambda}\right)}}{\sqrt{(1-\xi^2)(\lambda-a)(b-\lambda)}}\,g(\xi)\,d\xi,\qquad a<\lambda<b,$$

$$(3.2)\quad g(\xi)=\frac{1}{\pi}\sqrt{\frac{b-a}{2}}\operatorname*{l.i.m.}_{\delta\to 0,\,\delta>0}\int_{a+\delta}^{b-\delta}\frac{e^{-\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\lambda-a}{b-\lambda}\right)}}{\sqrt{(1-\xi^2)(\lambda-a)(b-\lambda)}}\,f(\lambda)\,d\lambda,\qquad -1<\xi<1,$$

are inversion formulae which represent isometries between the spaces $L^2(a,b)$ *and* $L^2(-1,1)$.

Proof. The Fourier inversion formulae are

$$F(v)=\frac{1}{\sqrt{2\pi}}\operatorname*{l.i.m.}_{N,M\to\infty}\int_{-N}^{M}e^{iuv}G(u)\,du,$$

$$G(u)=\frac{1}{\sqrt{2\pi}}\operatorname*{l.i.m.}_{N,M\to\infty}\int_{-N}^{M}e^{-iuv}F(v)\,dv.$$

We now introduce the new variables

$$u=\frac{1}{\sqrt{2\pi}}\log\frac{\xi+1}{1-\xi},\qquad v=\frac{1}{\sqrt{2\pi}}\log\frac{\lambda-a}{b-\lambda},$$

so that $\frac{du}{d\xi}=\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\frac{1}{1-\xi^2}$, $\frac{dv}{d\lambda}=\frac{b-a}{\sqrt{2\pi}}\frac{1}{(\lambda-a)(b-\lambda)}$. Next, we set $G(u)=\left(\frac{\pi}{2}\right)^{\frac{1}{2}}\sqrt{1-\xi^2(u)}\,G_1(u)$. Thus

$$\int_{-\infty}^{\infty}|G(u)|^2\,du=\int_{-1}^{1}|G_1(u(\xi))|^2\,d\xi.$$

If $G(u)$ belongs to $L^2(-\infty,\infty)$, then $F(v)$ belongs to $L^2(-\infty,\infty)$ and

$$\int_{-\infty}^{\infty}|F(v)|^2\,dv=\int_{-\infty}^{\infty}|G(u)|^2\,du.$$

We now set $F(v)=\frac{(2\pi)^{\frac{1}{2}}}{\sqrt{b-a}}\sqrt{(\lambda(v)-a)(b-\lambda(v))}\,F_1(v)$. Then

$$\int_{-\infty}^{\infty}|F(v)|^2\,dv=\int_{a}^{b}|F_1(v(\lambda))|^2\,d\lambda.$$

Furthermore,

$$F_1(v(\lambda)) = \frac{\sqrt{b-a}}{(2\pi)^{\frac{1}{2}}} \operatorname*{l.i.m.}_{\delta\to 0,\,\delta>0} \int_{-1+\delta}^{1-\delta} \frac{e^{\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\lambda-a}{b-\lambda}\right)}}{\left(\frac{\pi}{2}\right)^{\frac{1}{2}}\sqrt{(1-\xi^2)(\lambda-a)(b-\lambda)}} G_1(u(\xi))\,d\xi$$

$$= \frac{1}{\pi}\sqrt{\frac{b-a}{2}} \operatorname*{l.i.m.}_{\delta\to 0,\,\delta>0} \int_{-1+\delta}^{1-\delta} \frac{e^{\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\lambda-a}{b-\lambda}\right)}}{\sqrt{(1-\xi^2)(\lambda-a)(b-\lambda)}} G_1(u(\xi))\,d\xi .$$

Similarly,

$$G_1(u(\xi)) = \frac{1}{\pi}\sqrt{\frac{b-a}{2}} \operatorname*{l.i.m.}_{\delta\to 0,\,\delta>0} \int_{a+\delta}^{b-\delta} \frac{e^{-\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\lambda-a}{b-\lambda}\right)}}{\sqrt{(1-\xi^2)(\lambda-a)(b-\lambda)}} F_1(v(\lambda))\,d\lambda .$$

We are now in a position to prove

THEOREM 3.2. *The inversion formulae of Theorem 3.1 furnish a spectral representation for the finite Hilbert transformation $T_{(a,b)}$, i.e., if $f(\lambda)$ corresponds to $g(\xi)$, then $T_{(a,b)}f(\lambda)$ corresponds to $\xi g(\xi)$.*

Proof. The set of Hölder continuous functions on the interval $(-1, 1)$, which vanish identically near the points -1 and 1, is everywhere dense in the space $L^2(-1, 1)$. The set of images of these elements under the isometry of Theorem 3.1 is everywhere dense in $L^2(a, b)$. Let $g(\xi)$ be a function on the intervals $(-1, 1)$ which belongs to the set described above. Its image then has the form

$$f(\lambda) = \frac{1}{\pi}\sqrt{\frac{b-a}{2}} \int_{-1+\delta}^{1-\delta} \frac{e^{\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\lambda-a}{b-\lambda}\right)}}{\sqrt{(1-\xi^2)(\lambda-a)(b-\lambda)}} g(\xi)\,d\xi ,$$

where δ is sufficiently small. We now apply the operator $T_{(a,b)}$ to $f(\lambda)$:

$$\frac{1}{\pi i}\int_a^b \frac{f(\mu)}{\mu-\lambda}\,d\mu = \frac{1}{\pi^2 i}\sqrt{\frac{b-a}{2}} \int_a^b \frac{1}{\mu-\lambda}\left\{\int_{-1+\delta}^{1-\delta} \frac{e^{\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\mu-a}{b-\mu}\right)}}{\sqrt{(1-\xi^2)(\mu-a)(b-\mu)}} g(\xi)\,d\xi\right\} d\mu .$$

Using a method of TRICOMI [12], we shall show that the orders of integration in the last expression can be interchanged. We let

$$H(\mu,\xi) = \frac{e^{\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\mu-a}{b-\mu}\right)}}{\sqrt{(1-\xi^2)(\mu-a)(b-\mu)}} g(\xi) .$$

Next, we let $\{\varepsilon_n\}$ be a sequence of positive numbers such that $\lim_{n\to\infty}\varepsilon_n = 0$. Setting

$$\Phi_n(\xi) = \int_a^{\lambda-\varepsilon_n} \frac{H(\mu,\xi)}{\mu-\lambda}\,d\lambda + \int_{\lambda+\varepsilon_n}^b \frac{H(\mu,\xi)}{\mu-\lambda}\,d\mu ,$$

we see that

$$\lim_{n\to\infty} \Phi_n(\xi) = \int_a^b \frac{H(\mu,\xi)}{\mu-\lambda}\,d\mu.$$

Thus

$$\int_{-1+\delta}^{1-\delta}\left\{\int_a^b \frac{H(\mu,\xi)}{\mu-\lambda}\,d\mu\right\}d\xi = \int_{-1+\delta}^{1-\delta}\lim_{n\to\infty}\Phi_n(\xi)\,d\xi.$$

Using FUBINI's theorem, we see that

$$\int_a^b \frac{1}{\mu-\lambda}\left\{\int_{-1+\delta}^{1-\delta} H(\mu,\xi)\,d\xi\right\}d\mu = \lim_{n\to\infty}\int_{-1+\delta}^{1-\delta}\Phi_n(\xi)\,d\xi.$$

Now

$$|\Phi_n(\xi)| \leqq \int_a^b \left|\frac{H(\mu,\xi)-H(\lambda,\xi)}{\mu-\lambda}\right| d\mu + \left|H(\lambda,\xi)\log\frac{\lambda-a}{\lambda}\right|.$$

For a fixed λ, $a<\lambda<b$, the term $\left|H(\lambda,\xi)\log\frac{b-\lambda}{\lambda-a}\right|$ is bounded on the interval $-1+\delta\leqq\xi\leqq 1-\delta$. With λ fixed as before, the integral $\int_a^b \left|\frac{H(\mu,\xi)-H(\lambda,\xi)}{\mu-\lambda}\right| d\mu$ exists for each ξ in the interval $-1+\delta\leqq\xi\leqq 1-\delta$. In fact this integral represents a bounded function of ξ over this interval, since we may write

$$\int_a^b \left|\frac{H(\mu,\xi)-H(\lambda,\xi)}{\mu-\lambda}\right| d\mu = \int_a^{\lambda-\varepsilon}\left|\frac{H(\mu,\xi)-H(\lambda,\xi)}{\mu-\lambda}\right| d\mu +$$

$$+\int_{\lambda+\varepsilon}^b \left|\frac{H(\mu,\xi)-H(\lambda,\xi)}{\mu-\lambda}\right| d\mu + \int_{\lambda-\varepsilon}^{\lambda+\varepsilon}\left|\frac{H(\mu,\xi)-H(\lambda,\xi)}{\mu-\lambda}\right| d\mu$$

with $\varepsilon>0$ and sufficiently small. Thus we can apply LEBESGUE's theorem to the sequence $\{\Phi_n(\xi)\}$ to obtain

$$\lim_{n\to\infty}\int_{-1+\delta}^{1-\delta}\Phi_n(\xi)\,d\xi = \int_{-1+\delta}^{1-\delta}\lim_{n\to\infty}\Phi_n(\xi)\,d\xi.$$

This proves that we can invert the orders of integration. Thus for our special function $g(\xi)$ we have

$$\frac{1}{\pi i}\int_a^b \frac{f(\mu)}{\mu-\lambda}\,d\mu = \frac{1}{\pi}\sqrt{\frac{b-a}{2}}\int_{-1+\delta}^{1-\delta}\left\{\frac{1}{\pi i}\int_a^b \frac{e^{\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\mu-a}{b-\mu}\right)}}{\sqrt{(1-\xi^2)(\mu-a)(b-\mu)}}\,\frac{d\mu}{\mu-\lambda}\right\} g(\xi)\,d\xi$$

$$= \frac{1}{\pi}\sqrt{\frac{b-a}{2}}\int_{-1+\delta}^{1-\delta}\frac{e^{\frac{i}{2\pi}\left(\log\frac{\xi+1}{1-\xi}\right)\left(\log\frac{\lambda-a}{b-\lambda}\right)}}{\sqrt{(1-\xi^2)(\lambda-a)(b-\lambda)}}\,\xi\,g(\xi)\,d\xi,$$

from Lemma 2.1.

This proves Theorem 3.2 for functions $g(\xi)$ of our special class. Now let $g(\xi)$ be any function of the class $L^2(-1, 1)$ and let $f(\lambda)$ be its image under the isometry (3.1). Let $\{g_n(\xi)\}$ be a sequence of functions belonging to our special class, such that $\lim_{n\to\infty} g_n(\xi) = g(\xi)$ in the norm of $L^2(-1, 1)$. The corresponding images $f_n(\lambda)$ of the functions $g_n(\xi)$ under our isometry will converge to $f(\lambda)$ in the norm of the space $L^2(a, b)$. Because our ξ-interval is bounded, it is easy to see that the functions $\xi\, g_n(\xi)$ converge to $\xi g(\xi)$ in the norm of $L^2(-1, 1)$. Since $g_n(\xi)$ belongs to our special class, its image under our isometry is just $T_{(a,b)} f_n(\lambda)$, as we saw above. But from Lemma 1.1 we know that the operator $T_{(a,b)}$ is continuous in the norm of $L^2(a, b)$. Hence $\lim_{n\to\infty} T_{(a,b)} f_n(\lambda) = T_{(a,b)} f(\lambda)$. Let $h(\lambda)$ be the image of $g(\xi)$ under our isometry. Then

$$\begin{aligned}\|h(\lambda) - T_{(a,b)} f(\lambda)\|_{L^2(a,b)} &\leqq \|h(\lambda) - T_{(a,b)} f_n(\lambda)\|_{L^2(a,b)} + \|T_{(a,b)} f_n(\lambda) - T_{(a,b)} f(\lambda)\|_{L^2(a,b)} \\ &= \|\xi g(\xi) - \xi g_n(\xi)\|_{L^2(-1,1)} + \|T_{(a,b)} f_n(\lambda) - T_{(a,b)} f(\lambda)\|_{L^2(a,b)}.\end{aligned}$$

Letting $n \to \infty$, we see that $\|h(\lambda) - T_{(a,b)} f(\lambda)\|_{L^2(a,b)} = 0$, i.e. $h(\lambda) = T_{(a,b)} f(\lambda)$ almost everywhere on the interval (a, b). Thus the image of $\xi g(\xi)$ under our isometry is just $T_{(a,b)} f(\lambda)$. This completes the proof of the theorem.

4. Semi-infinite Hilbert transformations. We shall show that a semi-infinite Hilbert transformation is isometrically equivalent to a finite Hilbert transformation. Suppose that our operator is of the form

$$T_{(a,\infty)} f(\lambda) \equiv \frac{1}{\pi i} \int_a^\infty \frac{f(\mu)}{\mu - \lambda}\, d\mu.$$

Then we introduce the following change of variables:

We set $\hat{\mu} = \dfrac{\mu - a}{\mu - c}$, $\hat{\lambda} = \dfrac{\lambda - a}{\lambda - c}$, where $c < a$. Then $\mu = \dfrac{a - \hat{\mu} c}{1 - \hat{\mu}}$ and $\dfrac{d\mu}{d\hat{\mu}} = \dfrac{a - c}{(1 - \hat{\mu})^2}$, so that

$$\left\{\begin{aligned}\frac{1}{\pi i} \int_a^\infty \frac{f(\mu)}{\mu - \lambda}\, d\mu &= \frac{1}{\pi i} \int_0^1 \frac{f(\mu(\hat{\mu}))}{\dfrac{a - \hat{\mu} c}{1 - \hat{\mu}} - \dfrac{a - \hat{\lambda} c}{1 - \hat{\lambda}}} \cdot \frac{a - c}{(1 - \hat{\mu})^2}\, d\hat{\mu} \\ &= \frac{1 - \hat{\lambda}}{\pi i} \int_0^1 \frac{f(\mu(\hat{\mu}))/(1 - \hat{\mu})}{\hat{\mu} - \hat{\lambda}}\, d\hat{\mu}\ [\delta].\end{aligned}\right. \tag{4.1}$$

We now observe that the linear operator

$$V f(\lambda) \equiv \sqrt{a - c}\, \frac{f(\lambda(\hat{\lambda}))}{1 - \hat{\lambda}} \tag{4.2}$$

is an isometry of $L^2(a, \infty)$ onto $L^2(0, 1)$, since

$$\int_a^\infty |f(\lambda)|^2\, d\lambda = \int_0^1 \left| f(\lambda(\hat{\lambda})) \sqrt{\frac{d\lambda}{d\hat{\lambda}}} \right|^2 d\hat{\lambda} = \int_0^1 \left| \frac{\sqrt{a - c}\, f(\lambda(\hat{\lambda}))}{1 - \hat{\lambda}} \right|^2 d\hat{\lambda}.$$

We now have $T_{(0,1)} F(\hat{\lambda}) \equiv \frac{1}{\pi i} \int_0^1 \frac{F(\hat{\mu})}{\hat{\mu} - \hat{\lambda}} d\hat{\mu}$. From formulae (4.1) and (4.2) we now conclude that $T_{(0,1)} V = V T_{(a,\infty)}$ and therefore $T_{(0,1)} = V T_{(a,\infty)} V^{-1}$. Applying the isometries of Theorem 3.1 to $T_{(0,1)}$, we then obtain a spectral representation for the operator $T_{(a,\infty)}$.

Similar techniques can be used to analyze semi-infinite Hilbert transforms of the form $T_{(-\infty,b)}$.

5. A RELATION BETWEEN FINITE HILBERT TRANSFORMATIONS AND DIFFERENTIAL OPERATORS. Clearly the generalized eigenfunction of the finite Hilbert transformation (see Lemma 2.1) is a solution of the differential equation

$$i \sqrt{(\lambda - a)(b - \lambda)} \frac{d}{d\lambda} \sqrt{(\lambda - a)(b - \lambda)} f(\lambda) = -\left(\frac{b-a}{2\pi} \log \frac{\xi + 1}{1 - \xi}\right) f(\lambda).$$

We therefore consider the differential operator

$$D_{(a,b)} = i \sqrt{(\lambda - a)(b - \lambda)} \frac{d}{d\lambda} \sqrt{(\lambda - a)(b - \lambda)}$$

defined for functions $g(\lambda)$ of class $L^2(a, b)$, where $\sqrt{(\lambda - a)(b - \lambda)}\, g(\lambda)$ is absolutely continuous and the condition

$$\int_a^b \left| \sqrt{(\lambda - a)(b - \lambda)} \frac{d}{d\lambda} \sqrt{(\lambda - a)(b - \lambda)}\, g(\lambda) \right|^2 d\lambda < \infty$$

holds. The operator $D_{(a,b)}$ is then self-adjoint. In fact, a simple logarithmic change of variables of the type employed in the proof of Theorem 3.1 allows one to show that $D_{(a,b)}$ is isometrically equivalent to the simple self-adjoint differentiation operator $i\, d/du$ (with the usual domain of definition) defined for functions of $L^2(-\infty, \infty)$. The equivalence of this latter operator to a multiplication operator through Fourier transformations [1] and another logarithmic change of variables, permits one to draw the following conclusion:

THEOREM 5.1. *Under the isometry of Theorem 3.1, the differential operator $D_{(a,b)}$ is isometrically equivalent to the multiplication operator*

$$Q g(\xi) \equiv -\left(\frac{b-a}{2\pi} \log \frac{\xi + 1}{1 - \xi}\right) g(\xi)$$

defined for all $g(\xi)$ of class $L^2(-1, 1)$, such that

$$\int_{-1}^{1} \left| \left(\log \frac{\xi + 1}{1 - \xi}\right) g(\xi) \right|^2 d\xi < \infty.$$

Under the usual definitions of functions of operators we may then write

$$D_{(a,b)} = -\frac{b-a}{2\pi} \log \frac{T_{(a,b)} + 1}{1 - T_{(a,b)}}.$$

Bibliography

[1] ACHIESER, N. I., and I. M. GLASMANN: Theorie der linearen Operatoren im Hilbert-Raum. Berlin: Akademie-Verlag 1954. — [2] CARLEMAN, T.: Sur la resolution de certaines equations integrales. Ark. Mat., Astronom. Fys. **16** (1922). — [3] ELLIOTT, J.: On a class of integral equations. Proc. Amer. Math. Soc. **3**, 566—572 (1952). — [4] HAMEL, G.: Integralgleichungen. Berlin: Springer 1949. — [5] MICHLIN, S. G.: Singular integral equations. Amer. Math. Soc. Trans. No. 24, 1950. — [6] MUSKHELISHVILI, N. I.: Singular integral equations. Groningen: Noordhoff 1953. — [7] REISSNER, E.: Boundary value problems in aerodynamics of lifting surfaces in non-uniform motion. Bull. Amer. Math. Soc. **55**, 825—850 (1949). — [8] SÖHNGEN, H.: Zur Theorie der endlichen Hilbert-Transformation. Math. Z. **60**, 31—51 (1954). — [9] TITCHMARSH, E. C.: Introduction to the theory of Fourier integrals. Oxford 1937. — [10] TRICOMI, F. G.: On the finite Hilbert transformation. Quart. J. Math. (2) **2**, 199—211 (1951). — [11] TRICOMI, F. G.: Integral equations. New York: Interscience 1957. — [12] TRICOMI, F. G.: Sull'inversione dell'ordine di integrali "principali" nel senso di Cauchy. R. C. Accad. naz. Lincei (8) **18**, 3—7 (1955).

Yale University and New York University

(Eingegangen am 26. Januar 1959)

SALZMANN, H.
Math. Zeitschr. 71, 408–413 (1959)

Topologische Struktur zweidimensionaler projektiver Ebenen

Von

HELMUT SALZMANN

Bei der Untersuchung topologischer projektiver Ebenen spielen bei ganz verschiedenen Fragestellungen immer wieder diejenigen Ebenen eine ausgezeichnete Rolle, die topologisch homöomorph zur gewöhnlichen reellen projektiven Ebene sind (vgl. SALZMANN [23]–[25]). Das liegt einerseits an der besonders engen Abhängigkeit, die in solchen Ebenen zwischen verschiedenen Schließungssätzen besteht, andererseits auch daran, daß man ein im Grundgedanken auf Hilbert zurückgehendes allgemeines Verfahren kennt, eine Fülle sehr verschiedenartiger Beispiele solcher Ebenen zu konstruieren (vgl. HILBERT [9] § 23, MOULTON [17], MOHRMANN [15], TSCHETWERUCHIN [32], PIKKERT [19], KUIPER [12], ANDRÉ [3], [4], SALZMANN [23], [24]). Wesentlich für die Sonderstellung dieser Ebenenklasse ist die Tatsache (SALZMANN [23] § 10), daß ihre affinen Geraden homöomorph zur reellen Zahlengeraden sind (vgl. auch SKORNJAKOV [29], [30], [31]) und damit eine natürliche Anordnung tragen. Hier sollen nun für die zur reellen Ebene homöomorphen Ebenen schon früher (SALZMANN [24], [25]) teilweise angekündigte allgemeinere Kennzeichnungen gegeben werden. Sie sind in den folgenden beiden Sätzen enthalten.

Satz P. *Die folgenden Eigenschaften einer topologischen projektiven Ebene sind gleichwertig:*

(1) *Die Ebene ist lokal kompakt und topologisch zweidimensional.*

(2) *Eine ihrer Geraden ist lokal kompakt und topologisch eindimensional.*

(3) *Jede ihrer Geraden ist homöomorph zur reellen Kreislinie.*

(4) *Die Ebene ist homöomorph zur reellen projektiven Ebene.*

(5) *Die Ebene ist angeordnet und lokal kompakt in der Ordnungstopologie.*

Vielfach weiß man nur etwas über die Struktur der durch Weglassung einer Geraden entstehenden affinen Ebene. Darum werden dem Satz P affine Charakterisierungen der betrachteten Ebenen an die Seite gestellt.

Satz A. *Die folgenden Eigenschaften einer topologischen projektiven Ebene sind gleichwertig:*

(0) *Die projektive Ebene ist homöomorph zur reellen projektiven Ebene.*

(1) *Eine zugehörige affine Ebene ist lokal kompakt und topologisch zweidimensional.*

(2) *Eine affine Gerade ist lokal kompakt und topologisch eindimensional.*

(3) *Jede affine Gerade ist homöomorph zur reellen Zahlengeraden.*

(4) *Die affine Ebene ist homöomorph zur euklidischen Ebene.*

(5) *Die affine Ebene ist angeordnet und lokal kompakt in der Ordnungstopologie.*

Insbesondere folgt aus jeder der angeführten Eigenschaften die Kompaktheit der projektiven Ebene.

Eine topologische Ebene wird wie bei PICKERT [*18*] S. 261 und SKORNJAKOV [*28*] erklärt als eine projektive Ebene mit einer Topologie für die Punktmenge und die Geradenmenge, bei der Verbinden und Schneiden stetige Operationen sind. Jede Punktreihe ist dann homöomorph zu jedem Geradenbüschel, jede Gerade ist eine homogene, abgeschlossene Punktmenge, eine affine Ebene ist topologisches Produkt zweier affiner Geraden (s. PICKERT [*18*] S. 263—264). Eine topologische Eigenschaft soll der Ebene genau dann zukommen, wenn der Raum aller Punkte diese Eigenschaft besitzt. Nach SALZMANN [*23*] § 3 ist jede lokal kompakte projektive Ebene regulär von abzählbarer Basis, also metrisierbar und separabel, so daß alle gebräuchlichen Definitionen für die topologische Dimension äquivalent sind (vgl. HUREWICZ-WALLMAN [*10*]).

Beim Beweis der Sätze ist es zweckmäßig, zwischen der projektiven und der affinen Betrachtungsweise abzuwechseln. Die beiden Sätze werden darum nebeneinanderher bewiesen, und zwar nach dem Schema

$$\begin{array}{ccccccc} A0 & \Rightarrow & A1 & A2 \Leftarrow A3 & \Rightarrow & A4 & \Rightarrow A5 \\ \| & & \Downarrow & \Updownarrow & \Uparrow & \Downarrow & \Downarrow \\ P4 & & P1 & \Rightarrow P2 \Rightarrow P3 & & P4 & P5 \Rightarrow P2. \end{array}$$

Aus (A1) folgt (P1); denn alle zu einer projektiven Ebene gehörenden affinen Ebenen sind untereinander homöomorph und jede Gerade ist eine abgeschlossene Punktmenge, so daß jeder Punkt der projektiven Ebene innerer Punkt einer affinen Ebene ist. Daher übertragen sich die lokale Kompaktheit und die induktiv definierte Dimension (im Sinne von MENGER-URYSOHN) als lokale Eigenschaften von einer affinen Ebene auf die projektive.

Es sei jetzt eine lokal kompakte zweidimensionale projektive Ebene gegeben. Nach SALZMANN [*23*] § 5 ist dann die Ebene selbst und jede ihrer projektiven Geraden zusammenhängend, lokal zusammenhängend, kompakt und von abzählbarer Basis.

Ein topologischer Raum mit diesen Eigenschaften enthält ein topologisches Bild eines abgeschlossenen Intervalles reeller Zahlen (MAZURKIEWICZ [*14*], KURATOWSKI [*13*] S. 184). Wegen der Homogenität der projektiven Geraden hat die affine Gerade dieselbe Dimension wie die projektive Gerade. Daher hat die affine Ebene als topologisches Produkt zweier affiner Geraden mindestens die Dimension des topologischen Produkts einer projektiven Geraden mit einem reellen Intervall; also folgt nach ALEXANDROFF [*2*] (S. 236) aus der dimensionsmäßigen Vollwertigkeit des reellen Intervalls (vgl. auch MORITA [*16*], BOLTJANSKIJ [*5*]), daß die affine Ebene und erst recht die

projektive Ebene eine größere Dimension hat als die Gerade. Deswegen kann die Dimension einer Geraden einer (P1) erfüllenden Ebene nicht größer als 1 sein. Damit ist (P2) aus (P1) gefolgert. In der Bemerkung über die Dimensionsgleichheit der projektiven und der affinen Geraden ist gleichzeitig die Äquivalenz von (P2) mit (A2) enthalten.

Nun werde die Gültigkeit von (P2) vorausgesetzt. Dann ist wieder die Ebene und jede ihrer Geraden zusammenhängend, lokal zusammenhängend, kompakt und von abzählbarer Basis (SALZMANN [23] § 5). Nach der beim vorigen Beweisschritt gemachten Bemerkung enthält eine projektive Gerade also das topologische Bild eines abgeschlossenen reellen Intervalls, und da die Gruppe der Projektivitäten eine mindestens dreifach transitive (siehe z.B. PICKERT [18] S. 9) Homöomorphismengruppe der projektiven Geraden liefert, lassen sich in der affinen Geraden je zwei Punkte durch eine (zu einem reellen Intervall homöomorphe) einfache Kurve verbinden: Die affine Gerade ist kurvenweise zusammenhängend. Angenommen nun, die affine Gerade enthielte eine zur reellen Kreislinie (1-Sphäre) homöomorphe Teilmenge (Jordankurve) K. Wegen der Kompaktheit der Ebene kann man einen Satz von FREUDENTHAL [7] anwenden (s. auch SALZMANN [23] § 10), der besagt, daß sich in der affinen Geraden einer kompakten Ebene jede in der projektiven Geraden abgeschlossene Teilmenge durch eine stetige Schar von Homöomorphismen auf einen Punkt zusammenziehen läßt, der also insbesondere zeigt, daß jede stetige Abbildung der reellen Kreislinie in die affine Gerade homotop zu einer Konstanten ist. Das bedeutet, daß es eine stetige Abbildung des Zylindermantels in die affine Gerade gibt, die dessen eine Randlinie homöomorph auf die Jordankurve K abbildet, während das Bild der anderen Randlinie aus einem Punkt besteht. Das Bild des Zylindermantels ist eine kompakte Teilmenge T der eindimensionalen affinen Geraden, in der die Zusammenziehung von K also auch schon möglich ist. K erzeugt in der Čechschen Homologie dieser Teilmenge T einen eindimensionalen Zyklus, der homolog 0 ist. Andererseits kann in einem eindimensionalen Raum kein eindimensionaler Zyklus beranden, womit die Annahme der Existenz der Jordankurve K auf einen Widerspruch geführt hat. Anders: Als kompakter Raum mit abzählbarer Basis ist T metrisierbar. Nach ALEXANDROFF [1] ist ein kompakter metrischer Raum genau dann eindimensional, wenn es zu jedem $\varepsilon > 0$ eine stetige ε-Abbildung (d.h. eine stetige Abbildung, bei der das Urbild jedes Punktes einen Durchmesser $< \varepsilon$ hat) auf einen eindimensionalen endlichen Komplex, nämlich den Nerv einer ε-Überdeckung, gibt (vgl. HUREWICZ-WALLMAN [10] S. 73, PONTRJAGIN [20] Satz 4). Für hinreichend kleines ε wird dabei K auf einen wesentlichen Zyklus des Streckenkomplexes abgebildet, der in dem Komplex zusammenziehbar sein müßte, was unmöglich ist. Damit ist unter der Voraussetzung von (P2) bewiesen, daß die affinen Geraden kurvenweise zusammenhängend sind und keine geschlossenen einfachen Kurven enthalten. Dann zeigt ein früher (SALZMANN [23] § 10) unter der Annahme von (P4) geführter Beweis, der aber nur genau die beiden eben genannten Eigenschaften der Geraden benutzt, die Homöomorphie der projektiven Geraden zur reellen Kreislinie und damit (P3).

Die Eigenschaft (A3) ist eine unmittelbare Folge von (P3) und schließt insbesondere (A2) in sich ein.

Der Schluß von (A3) auf (A4) ist ein Spezialfall der schon zitierten Darstellbarkeit der affinen Ebene als topologisches Produkt zweier affiner Geraden (SALZMANN [22]).

Ist nun die affine Ebene homöomorph zur euklidischen Ebene, so sind insbesondere (A1) und alle schon daraus gezogenen Folgerungen erfüllt. Die projektive Ebene ist dann als lokal euklidischer topologischer Raum eine zweidimensionale kompakte Mannigfaltigkeit und wegen der Triangulierbarkeit der zweidimensionalen Mannigfaltigkeiten (s. RADÓ [21], GAWEHN [8]) eine geschlossene Fläche im Sinne der kombinatorischen Topologie (vgl. SEIFERT-THRELFALL [27]). Die topologische Struktur der geschlossenen Flächen ist wohlbekannt (Hauptsatz der Flächentopologie, s. KERÉKJÁRTÓ [11] S. 149, SEIFERT-THRELFALL [27] S. 141). Wegen (P3) ist jede projektive Gerade eine Jordankurve in der Fläche und die nach Herausschneiden der Geraden übrigbleibende affine Ebene ist nach Voraussetzung homöomorph zu einer offenen Kreisscheibe. Daher läßt sich jede (nicht 0-homotope) Jordankurve in der geschlossenen Fläche durch eine stetige Deformation in eine feste projektive Gerade überführen (jede echte Teilmenge der Geraden ist 0-homotop). Die Fundamentalgruppe der projektiven Ebene wird also von einer (und damit von jeder) projektiven Geraden erzeugt, d.h. sie ist zyklisch. Die einzigen Flächen mit zyklischer Fundamentalgruppe sind aber die Kugel und die zur reellen projektiven Ebene homöomorphe nichtorientierbare Fläche vom Geschlecht 1 (s. SEIFERT-THRELFALL [27] S. 170). Die Kugel wird im Gegensatz zur projektiven Ebene von jeder Jordankurve in zwei offene Kreisscheiben zerlegt (Jordanscher Kurvensatz, s. z.B. E. SCHMIDT [26]). Damit ist (P4) = (A0) aus (A4) hergeleitet.

Wegen der Abgeschlossenheit der Geraden folgt aus (A0) wieder (A1) und damit die Gleichwertigkeit aller ohne Anordnung formulierten Eigenschaften.

Es sei jetzt die affine Ebene homöomorph zur euklidischen Ebene. Dann ist nach dem eben Bemerkten auch (A3) erfüllt. Durch die Anordnung der reellen Zahlen wird also in jeder affinen Geraden in natürlicher Weise eine Zwischenbeziehung hervorgerufen (vgl. PICKERT [18] § 9.1), und zwar zerlegt jeder Punkt einer affinen Geraden die Gerade in zwei zusammenhängende Komponenten, so daß er genau zwischen jedem Paar von Punkten aus verschiedenen Komponenten der Geraden liegt. Wegen der Darstellbarkeit der affinen Ebene als topologisches Produkt zweier beliebiger nicht paralleler Geraden folgt aus dieser Kennzeichnung der Zwischenbeziehung ihre Invarianz gegenüber Parallelprojektionen. Nach CRAMPE [6] § 4 ist die Invarianz der Zwischenbeziehung gegenüber Parallelprojektionen bereits hinreichend für ihre Invarianz gegenüber beliebigen affinen Projektivitäten, d.h. dafür, daß die Zwischenbeziehung eine Anordnung der affinen Ebene definiert. Die Ordnungstopologie wird nun dadurch gewonnen, daß man die von den offenen Intervallen auf den affinen Geraden erzeugte lineare Topologie durch Produktbildung zweier Geraden auf die affine Ebene überträgt (vgl. WYLER [33]).

Sie stimmt also im vorliegenden Fall mit der gegebenen Topologie überein. Insbesondere ist die affine Ebene lokal kompakt in der Ordnungstopologie, d.h. sie erfüllt (A5).

Die Anordnung einer affinen Ebene läßt sich eindeutig zu einer Anordnung der projektiven Ebene fortsetzen (s. PICKERT [18] § 9.2). Die Ordnungstopologie der projektiven Ebene stimmt lokal mit der Ordnungstopologie der affinen Ebene überein. Also ist die projektive Ebene genau dann lokal kompakt in der Ordnungstopologie, wenn die affine Ebene es ist. Daher ist (P5) äquivalent mit (A5).

Schließlich sei eine angeordnete lokal kompakte projektive Ebene gegeben. Dann ist auf jeder ihrer affinen Geraden eine Zwischenbeziehung erklärt. Wählt man eine der beiden durch die Zwischenbeziehung hervorgerufenen Anordnungen (s. PICKERT [18] § 9.1) einer affinen Geraden K aus, so ist K eine linear geordnete in der Ordnungstopologie lokal kompakte Menge. Weiter ist K dicht geordnet, d.h. zwischen je zwei Punkten liegen weitere Punkte von K (s. z.B. CRAMPE [6] Hilfsatz 1). Wegen der lokalen Kompaktheit gibt es in K ein offenes Intervall (a, b) mit kompakter Abschließung. Die Abschließung von (a, b) ist das abgeschlossene Intervall $[a, b]$, da K dicht geordnet ist, also a, b Häufungspunkte von (a, b) sind. Es soll nun gezeigt werden, daß $[a, b]$ positive Dimension hat. Wäre nämlich $[a, b]$ nulldimensional, so hätte etwa a beliebig kleine randlose (d.h. offen und abgeschlossene) Umgebungen in $[a, b]$. Sei U eine solche Umgebung, die b nicht enthält. Weil das Komplement von U in $[a, b]$ als abgeschlossene Teilmenge von $[a, b]$ kompakt ist, gibt es einen ersten Punkt $c \in [a, b]$, der nicht zu U gehört. Das Intervall $[a, c)$ ist offen in $[a, b]$ und in U enthalten. Da $[a, b]$ dicht geordnet ist, gehört c zur Abschließung von $[a, c)$; aber diese liegt in U, weil U abgeschlossen ist. Das widerspricht der Konstruktion von c. Dieser Widerspruch zeigt, daß $[a, b]$ und damit K nicht nulldimensional ist. Jeder Punkt von K hat eine Umgebungsbasis aus offenen Intervallen. Der Rand eines offenen Intervalls kann höchstens die beiden Endpunkte enthalten, ist also in jedem Fall nulldimensional. Daher kann die Dimension von K nicht größer als 1 sein. Es folgt, daß jede Gerade lokal kompakt und eindimensional ist. Damit sind die beiden Sätze vollständig bewiesen.

Bemerkung. Mit Satz P ist das von FREUDENTHAL [7] erwähnte Problem, ob die Geraden einer kompakten zusammenhängenden projektiven Ebene homöomorph zu Sphären sind, im Fall der niedrigsten Dimension gelöst.

Literatur

[1] ALEXANDROFF, P.: Über den allgemeinen Dimensionsbegriff und seine Beziehungen zur elementaren geometrischen Anschauung. Math. Ann. **98**, 617–636 (1928). — [2] ALEXANDROFF, P.: Dimensionstheorie. Ein Beitrag zur Geometrie der abgeschlossenen Mengen. Math. Ann. **106**, 161–238 (1932). — [3] ANDRÉ, J.: Affine Ebenen mit genügend vielen Translationen. Math. Nachr. **19**, 203–210 (1958). — [4] ANDRÉ, J.: Homomorphismen projektiver Ebenen. Erscheint demnächst. — [5] BOLTJANSKIJ, V.: Über die dimensionelle Vollwertigkeit der Kompakta. [Russisch]. Dokl. Akad. Nauk SSSR. **67**, 773–776 (1949). Zbl. Math. **37**, 98. — [6] CRAMPE, S.: Angeordnete projektive Ebenen. Math. Z.

69, 435–462 (1958). — [7] FREUDENTHAL, H.: Kompakte projektive Ebenen. Illinois J. Math. **1**, 9–13 (1957). — [8] GAWEHN, I.: Über unberandete 2-dimensionale Mannigfaltigkeiten. Math. Ann. **98**, 321–354 (1928). — [9] HILBERT, D.: Grundlagen der Geometrie, 4. Aufl. Leipzig 1913. — [10] HUREWICZ-WALLMAN: Dimension theory. Princeton 1948. — [11] KERÉKJÁRTÓ, B. v.: Vorlesungen über Topologie. I. Flächentopologie. Berlin 1923. — [12] KUIPER, N. H.: A real analytic non-desargesian plane. Nieuw Arch. Wiskunde, III. R. **5**, 19–24 (1957). — [13] KURATOWSKI, C.: Topologie, II. 2. Aufl. Warschau 1952. — [14] MAZURKIEWICZ, S.: Sur les lignes de Jordan. Fund. Math. **1**, 166–209 (1920). Fortschr. d. Math. **47**, 521. — [15] MOHRMANN, H.: Hilbertsche und Beltramische Liniensysteme. Math. Ann. **85**, 177–183 (1922). Fortschr. d. Math. **48**, 638. — [16] MORITA, K.: On the dimension of product spaces. Amer. J. Math. **75**, 205–223 (1953). — [17] MOULTON, F. R.: A simple non-desargesian plane geometry. Trans. Amer. Math. Soc. **3**, 192–195 (1902). Fortschr. d. Math. **33**, 497. — [18] PICKERT, G.: Projektive Ebenen. Berlin-Göttingen-Heidelberg 1955. — [19] PICKERT, G.: Eine nichtdesargessche Ebene mit einem Körper als Koordinatenbereich. Publ. Math. Debrecen **4**, 157–160 (1956). Zbl. Math. **70**, 156. — [20] PONTRJAGIN, L. S.: Grundzüge der kombinatorischen Topologie. Moskau-Leningrad 1947, deutsch Berlin 1956. — [21] RADÓ, T.: Über den Begriff der Riemannschen Fläche. Acta Sci. Math. Szeged **2**, 101–121 (1925). Fortschr. d. Math. **51**, 273. — [22] SALZMANN, H.: Über den Zusammenhang in topologischen projektiven Ebenen. Math. Z. **61**, 489–494 (1955). — [23] SALZMANN, H.: Topologische projektive Ebenen. Math. Z. **67**, 436–466 (1957). — [24] SALZMANN, H.: Kompakte zweidimensionale projektive Ebenen. Arch. Math. **9**, 447–454 (1958). — [25] SALZMANN, H.: Homomorphismen topologischer projektiver Ebenen. Arch. Math. **10**, 51–55 (1959). — [26] SCHMIDT, E.: Über den Jordanschen Kurvensatz. Sitzgsber. preuß. Akad. Wiss. Berlin **1923**, 318–329. Fortschr. d. Math. **49**, 144. — [27] SEIFERT-THRELFALL: Lehrbuch der Topologie. Leipzig 1934. — [28] SKORNJAKOV, L. A.: Topologische projektive Ebenen. [Russisch]. Trudy Moskov. Mat. Obšč. **3**, 347–373 (1954). Zbl. Math. **57**, 362. — [29] SKORNJAKOV, L. A.: Kurvensysteme in der Ebene. [Russisch]. Dokl. Akad. Nauk SSSR. **98**, 25–26 (1954). Zbl. Math. **56**, 162. — [30] SKORNJAKOV, L. A.: Die Metrisation der projektiven Ebene im Zusammenhang mit einem gegebenen Kurvensystem. [Russisch]. Izv. Akad. Nauk SSSR., Ser. Mat. **19**, 471–482 (1955). Zbl. Math. **67**, 401. — [31] SKORNJAKOV, L. A.: Kurvensysteme in der Ebene. [Russisch]. Trudy Moskov. Mat. Obšč. **6**, 135–164 (1957). — [32] TSCHETWERUCHIN, N. F.: Eine Bemerkung zu den Nicht-Desarguesschen Liniensystemen. Jber. dtsch. Math.-Ver. **36**, 134–136 (1927). — [33] WYLER, O.: Order and topology in projective planes. Amer. J. Math. **74**, 656–666 (1952).

Frankfurt a.M., Mathematisches Seminar der Universität

(Eingegangen am 30. Januar 1959)

ERWE, F.
Math. Zeitschr. 71, 414–426 (1959)

Abschätzungen für beschränkte Funktionensysteme

Herrn WOLFGANG KRULL zum 60. Geburtstag am 26. 8. 59 gewidmet

Von

FRIEDHELM ERWE

Die Abschätzungen über beschränkte Funktionensysteme, wie sie in einer früheren Arbeit von E. PESCHL und dem Verfasser gegeben wurden[1]), lassen sich in mannigfacher Weise abwandeln und verallgemeinern. Die in der vorliegenden Arbeit niedergelegten Untersuchungen erstrecken sich insbesondere auf Abschätzungen für die Differenz zweier beschränkter Funktionensysteme. Es ergeben sich dabei Aussagen über Kreisscheiben, in denen zwei beschränkte Funktionensysteme, für die die Anfangskoeffizienten ihrer Potenzreihenentwicklung um den Nullpunkt fest vorgegeben sind, in ihren Werten nicht übereinstimmen können.

In Bezeichnungen und Schreibweise schließen wir uns im wesentlichen an die genannte frühere Arbeit[1]) an. Ein System aus n Funktionen $f_1(z)$, $f_2(z), \ldots, f_n(z)$ wird als n-gliedrige Spalte (Funktionenspalte) geschrieben:

$$\mathfrak{f}(z) = \begin{pmatrix} f_1(z) \\ f_2(z) \\ \vdots \\ f_n(z) \end{pmatrix},$$

der Absolutbetrag einer solchen Spalte durch

$$|\mathfrak{f}(z)| = +\sqrt{|f_1(z)|^2 + |f_2(z)|^2 + \cdots + |f_n(z)|^2}$$

definiert. Der Übergang zur konjugiert komplexen und zugleich transponierten Spalte (allgemeiner: Matrix) werde durch * gekennzeichnet. Im Matrizenkalkül ist dann

$$|\mathfrak{f}(z)|^2 = \mathfrak{f}^*(z)\,\mathfrak{f}(z).$$

Eine Funktionenspalte $\mathfrak{f}(z)$ heißt *beschränkt*, wenn sie im Einheitskreis $|z|<1$ holomorph ist (Holomorphie aller Glieder von $\mathfrak{f}$ ist gemeint), und wenn in $|z|<1$

$$|\mathfrak{f}(z)| \leqq 1 \tag{1}$$

gilt.

Grundlegend ist das Schwarzsche Lemma: *Wenn* $\mathfrak{f}(z)$ *beschränkt ist, und wenn* $\mathfrak{f}(0) = \mathfrak{o}$ *(Nullspalte) ist, so gilt:*

$$|\mathfrak{f}(z)| \leqq |z| \quad \text{für} \quad |z|<1. \tag{2}$$

[1]) Über beschränkte Systeme von Funktionen. Math. Ann. **126**, 185–220 (1953).

Gilt hier für ein $z \neq 0$ *das Gleichheitszeichen, so ist*

$$\mathfrak{f}(z) \equiv \mathfrak{c}\, z \quad \textit{für} \quad |z| < 1, \tag{3}$$

wo $\mathfrak{c}$ *eine konstante Spalte vom Absolutbetrag* 1 *ist.*

Dieser Satz wurde von E. PESCHL und Verfasser[1]) über den allgemeinen Satz vom Maximum für Funktionenspalten bewiesen. Bei BOCHNER-MARTIN[2]) wird er auf den klassischen Fall $n = 1$ zurückgeführt (hier und im folgenden soll n stets die Gliederzahl der vorkommenden Spalten bedeuten), und zwar vermöge der folgenden offenbar richtigen Aussage: *Eine Funktionenspalte* $\mathfrak{f}(z)$ *ist genau dann beschränkt, wenn für jede konstante Spalte* $\mathfrak{c}$ *vom Absolutbetrag* 1 *die Funktion* $\mathfrak{c}^* \mathfrak{f}(z)$ *beschränkt ist.*

Diese notwendige und hinreichende Bedingung mag oft von Nutzen sein. Die Aussage über die Gültigkeit des Gleichheitszeichens in (2) wird aber aus ihr nicht unmittelbar erkannt. Sie ist auch keineswegs geeignet, die folgenden Abschätzungen unmittelbar auf entsprechende klassische zurückzuführen. Meines Wissens sind sogar die meisten dieser Abschätzungen auch für den klassischen Fall $n = 1$ noch nicht gemacht worden. Besonders interessant sind übrigens die oft erheblichen Unterschiede, die sich für $n > 1$ gegenüber $n = 1$ in den Formeln ergeben.

Mittels der analytischen Automorphismen der Einheitshyperkugel $|\mathfrak{z}| < 1$ des komplex n-dimensionalen $\mathfrak{z}$-Raumes, nämlich der Abbildungen (von unitären Drehungen abgesehen)

$$\mathfrak{z} \leftrightarrow \Gamma(\mathfrak{a})\, \frac{\mathfrak{z} - \mathfrak{a}}{1 - \mathfrak{a}^* \mathfrak{z}} \qquad (|\mathfrak{z}| < 1),$$

$$|\mathfrak{a}| < 1, \qquad \Gamma(\mathfrak{a}) = \frac{1}{1 + v(\mathfrak{a})}\, \mathfrak{a}\, \mathfrak{a}^* + v(\mathfrak{a})\, \mathsf{E}, \qquad v(\mathfrak{a}) = + \sqrt{1 - |\mathfrak{a}|^2}$$

(E = Einheitsmatrix) ergibt sich aus (2) die Ungleichung

$$\left| \Gamma\big(\mathfrak{f}(z_0)\big)\, \frac{\mathfrak{f}(z) - \mathfrak{f}(z_0)}{1 - \mathfrak{f}^*(z_0)\, \mathfrak{f}(z)} \right| \leqq \left| \frac{z - z_0}{1 - \bar{z}_0 z} \right| \quad \text{für} \quad |z| < 1,\ |z_0| < 1 \ \big(|\mathfrak{f}(z_0)| < 1\big) \tag{4}$$

und eine entsprechende Aussage über die Gültigkeit des Gleichheitszeichens.

An Rechenregeln für die Matrizen Γ werden insbesondere benötigt:

$$\Gamma(\mathfrak{a}) = \mathsf{E} \quad \text{für} \quad n = 1,$$

$$\Gamma^*(\mathfrak{a}) = \Gamma(\mathfrak{a}),$$

$$\Gamma(s\,\mathfrak{a})\, \mathfrak{a} = \mathfrak{a}, \qquad \mathfrak{a}^* \Gamma(s\,\mathfrak{a}) = \mathfrak{a}^* \qquad (s \text{ komplex};\ |s\,\mathfrak{a}| < 1),$$

$$\Gamma^2(\mathfrak{a}) = v^2(\mathfrak{a})\, \mathsf{E} + \mathfrak{a}\, \mathfrak{a}^*,$$

$$\Gamma^{-2}(\mathfrak{a}) = \frac{1}{v^2(\mathfrak{a})}\, (\mathsf{E} - \mathfrak{a}\, \mathfrak{a}^*),$$

[2]) Several Complex Variables. Princeton 1948.

also

$$\Gamma^{-1}(\mathfrak{a}) = \frac{1}{v^2(\mathfrak{a})}\left(\Gamma(\mathfrak{a}) - \mathfrak{a}\,\mathfrak{a}^*\right).$$

Wenn $\mathfrak{a}$, $\mathfrak{b}$ linear abhängig über dem Körper der komplexen Zahlen sind, so ist

$$\Gamma(\mathfrak{a})\,\Gamma(\mathfrak{b}) = \Gamma(\mathfrak{b})\,\Gamma(\mathfrak{a}),$$

also auch

$$\Gamma^{-1}(\mathfrak{a})\,\Gamma(\mathfrak{b}) = \Gamma(\mathfrak{b})\,\Gamma^{-1}(\mathfrak{a}).$$

Hilfssatz. *Seien die Spalte $\mathfrak{a}$ und die komplexe Zahl s vorgegeben ($|\mathfrak{a}|<1$, $|s|<1$). Dann ist*

$$(5)\qquad \begin{cases} M(s,\mathfrak{a}) = \max\limits_{|\mathfrak{c}|\le 1}\left(\dfrac{1-|\mathfrak{a}|^2}{1-|\mathfrak{a}|^2|s|^2}\,|\Gamma^{-1}(\mathfrak{a})\,\Gamma(s\,\mathfrak{a})\,\mathfrak{c}|\right) \\ \qquad = \begin{cases} \dfrac{1-|\mathfrak{a}|^2}{1-|\mathfrak{a}|^2|s|^2} & \text{für} \quad n=1, \\ \sqrt{\dfrac{1-|\mathfrak{a}|^2}{1-|\mathfrak{a}|^2|s|^2}} & \text{für} \quad n>1. \end{cases} \end{cases}$$

Das Maximum wird genau für solche $\mathfrak{c}$ angenommen, für die gilt:

$$|\mathfrak{c}| = 1 \qquad (\text{im Falle } n=1),$$

$$|\mathfrak{c}| = 1, \quad \mathfrak{a}^*\mathfrak{c} = 0 \qquad (\text{im Falle } n>1).$$

Beweis. Der Fall $n=1$ ist trivial. Sei nun $n>1$. Es ist

$$\begin{aligned} |\Gamma^{-1}(\mathfrak{a})\,\Gamma(s\,\mathfrak{a})\,\mathfrak{c}|^2 &= \mathfrak{c}^*\Gamma^*(s\,\mathfrak{a})\,\Gamma^{*-1}(\mathfrak{a})\,\Gamma^{-1}(\mathfrak{a})\,\Gamma(s\,\mathfrak{a})\,\mathfrak{c} \\ &= \mathfrak{c}^*\Gamma^{-2}(\mathfrak{a})\,\Gamma^2(s\,\mathfrak{a})\,\mathfrak{c} \\ &= \frac{1}{v^2(\mathfrak{a})}\,\mathfrak{c}^*(\mathsf{E}-\mathfrak{a}\,\mathfrak{a}^*)\left(v^2(s\,\mathfrak{a})\,\mathsf{E}+\mathfrak{a}\,\mathfrak{a}^*\right)\mathfrak{c} \\ &= \frac{1}{v^2(\mathfrak{a})}\,\mathfrak{c}^*\left(v^2(s\,\mathfrak{a})\,\mathsf{E}-(1-|s|^2)\,\mathfrak{a}\,\mathfrak{a}^*\right)\mathfrak{c} \\ &= \frac{1}{v^2(\mathfrak{a})}\left(v^2(s\,\mathfrak{a})\,|\mathfrak{c}|^2-(1-|s|^2)\,|\mathfrak{a}^*\mathfrak{c}|^2\right) \\ &\le \frac{v^2(s\,\mathfrak{a})}{v^2(\mathfrak{a})}, \end{aligned}$$

und das Gleichheitszeichen steht offenbar genau dann, wenn $|\mathfrak{a}^*\mathfrak{c}|=0$ und $|\mathfrak{c}|=1$ ist. Damit ist alles bewiesen.

Mit den nun bereitgestellten Mitteln beweisen wir den folgenden

Satz 1. *Seien $\mathfrak{a}$, $\mathfrak{b}$ gegebene Spalten mit $|\mathfrak{a}|<1$, $|\mathfrak{b}|<1$. Für je zwei beschränkte Funktionenspalten $\mathfrak{f}(z)$, $\mathfrak{g}(z)$ mit $\mathfrak{f}(0)=\mathfrak{a}$, $\mathfrak{g}(0)=\mathfrak{b}$ gilt:*

$$(6)\quad |\mathfrak{f}(z)-\mathfrak{g}(z)| \le (1-|z|^2)\left|\frac{\mathfrak{a}}{1-|\mathfrak{a}|^2|z|^2}-\frac{\mathfrak{b}}{1-|\mathfrak{b}|^2|z|^2}\right| + |z|\left(M(z,\mathfrak{a})+M(z,\mathfrak{b})\right),$$

$$(7)\quad |\mathfrak{f}(z)-\mathfrak{g}(z)| \ge (1-|z|^2)\left|\frac{\mathfrak{a}}{1-|\mathfrak{a}|^2|z|^2}-\frac{\mathfrak{b}}{1-|\mathfrak{b}|^2|z|^2}\right| - |z|\left(M(z,\mathfrak{a})+M(z,\mathfrak{b})\right).$$

(6) ist für jedes z scharf, falls $n=1$ oder $\mathfrak{a}=\mathfrak{b}$ ist. Das gleiche gilt für (7), solange die rechte Seite nicht negativ wird.

Beweis. Es wird

$$\mathfrak{f}_1(z) = \Gamma(\mathfrak{a})\,\frac{\mathfrak{f}(z)-\mathfrak{a}}{1-\mathfrak{a}^*\mathfrak{f}(z)}\cdot\frac{1}{z} \tag{8}$$

gesetzt, und gemäß (4) ist $\mathfrak{f}_1$ beschränkt (nach stetiger Ergänzung im Nullpunkt). Es ist

$$\mathfrak{f} = \Gamma(\mathfrak{a})\,\frac{z\mathfrak{f}_1+\mathfrak{a}}{1+\mathfrak{a}^* z\mathfrak{f}_1} = z\,\Gamma(\mathfrak{a})\,\frac{\mathfrak{f}_1+\overline{z}\,\mathfrak{a}}{1+\mathfrak{a}^* z\mathfrak{f}_1} + (1-z\overline{z})\,\frac{\mathfrak{a}}{1+\mathfrak{a}^* z\mathfrak{f}_1}.$$

Nun führen wir

$$\mathfrak{f}_2 = \Gamma(\overline{z}\,\mathfrak{a})\,\frac{\mathfrak{f}_1+\overline{z}\,\mathfrak{a}}{1+\mathfrak{a}^* z\mathfrak{f}_1} \tag{9}$$

ein, und es ist offenbar $|\mathfrak{f}_2|\leqq 1$. Man beachte aber, daß $\mathfrak{f}_2=\mathfrak{f}_2(z)$ in $|z|<1$ nicht holomorph ist, jedenfalls für $\mathfrak{a}\neq 0$ nicht. Da

$$1+\mathfrak{a}^* z\mathfrak{f}_1 = \frac{1-|\mathfrak{a}|^2|z|^2}{1-\mathfrak{a}^* z\mathfrak{f}_2}$$

ist, folgt:

$$\begin{aligned}\mathfrak{f} &= z\,\Gamma(\mathfrak{a})\,\Gamma^{-1}(\overline{z}\,\mathfrak{a})\,\mathfrak{f}_2 + \frac{1-|z|^2}{1-|\mathfrak{a}|^2|z|^2}\,\mathfrak{a}\,(1-\mathfrak{a}^* z\,\mathfrak{f}_2)\\ &= \frac{1-|z|^2}{1-|\mathfrak{a}|^2|z|^2}\,\mathfrak{a} + z\left(\Gamma(\mathfrak{a})\,\Gamma^{-1}(\overline{z}\,\mathfrak{a}) - \frac{1-|z|^2}{1-|\mathfrak{a}|^2|z|^2}\,\mathfrak{a}\,\mathfrak{a}^*\right)\mathfrak{f}_2.\end{aligned}$$

Auf Grund obiger Rechenregeln für die Matrizen Γ ergibt sich

$$\begin{aligned}\Gamma(\mathfrak{a})\,\Gamma^{-1}(\overline{z}\,\mathfrak{a}) - \frac{1-|z|^2}{1-|\mathfrak{a}|^2|z|^2}\,\mathfrak{a}\,\mathfrak{a}^* &= \Gamma(\mathfrak{a})\left(\Gamma^{-1}(\overline{z}\,\mathfrak{a}) - \frac{1-|z|^2}{v^2(\overline{z}\,\mathfrak{a})}\,\mathfrak{a}\,\mathfrak{a}^*\right)\\ &= \frac{\Gamma(\mathfrak{a})}{v^2(\overline{z}\,\mathfrak{a})}\left(\Gamma(\overline{z}\,\mathfrak{a}) - \mathfrak{a}\,\mathfrak{a}^*\right)\\ &= \frac{\Gamma(\mathfrak{a})}{v^2(\overline{z}\,\mathfrak{a})}\left(\mathsf{E}-\mathfrak{a}\,\mathfrak{a}^*\right)\Gamma(\overline{z}\,\mathfrak{a})\\ &= \frac{v^2(\mathfrak{a})}{v^2(\overline{z}\,\mathfrak{a})}\,\Gamma^{-1}(\mathfrak{a})\,\Gamma(\overline{z}\,\mathfrak{a}).\end{aligned}$$

Wir haben somit

$$\mathfrak{f} = \frac{1-|z|^2}{1-|\mathfrak{a}|^2|z|^2}\,\mathfrak{a} + z\,\frac{1-|\mathfrak{a}|^2}{1-|\mathfrak{a}|^2|z|^2}\,\Gamma^{-1}(\mathfrak{a})\,\Gamma(\overline{z}\,\mathfrak{a})\,\mathfrak{f}_2. \tag{10}$$

Mit

$$\mathfrak{g}_1(z) = \Gamma(\mathfrak{b})\,\frac{\mathfrak{g}(z)-\mathfrak{b}}{1-\mathfrak{b}^*\mathfrak{g}(z)}\cdot\frac{1}{z}, \qquad \mathfrak{g}_2 = \Gamma(\overline{z}\,\mathfrak{b})\,\frac{\mathfrak{g}_1+\overline{z}\,\mathfrak{b}}{1+\mathfrak{b}^* z\,\mathfrak{g}_1} \tag{11}$$

folgt analog:

$$\mathfrak{g} = \frac{1-|z|^2}{1-|\mathfrak{b}|^2|z|^2}\,\mathfrak{b} + z\,\frac{1-|\mathfrak{b}|^2}{1-|\mathfrak{b}|^2|z|^2}\,\Gamma^{-1}(\mathfrak{b})\,\Gamma(\overline{z}\,\mathfrak{b})\,\mathfrak{g}_2. \tag{12}$$

Aus (10) und (12) ergibt sich durch Differenzbildung

$$\begin{aligned}\mathfrak{f}-\mathfrak{g} &= (1-|z|^2)\left(\frac{\mathfrak{a}}{1-|\mathfrak{a}|^2|z|^2} - \frac{\mathfrak{b}}{1-|\mathfrak{b}|^2|z|^2}\right) +\\ &\quad + z\left(\frac{1-|\mathfrak{a}|^2}{1-|\mathfrak{a}|^2|z|^2}\,\Gamma^{-1}(\mathfrak{a})\,\Gamma(\overline{z}\,\mathfrak{a})\,\mathfrak{f}_2 - \frac{1-|\mathfrak{b}|^2}{1-|\mathfrak{b}|^2|z|^2}\,\Gamma^{-1}(\mathfrak{b})\,\Gamma(\overline{z}\,\mathfrak{b})\,\mathfrak{g}_2\right),\end{aligned} \tag{13}$$

und hieraus folgen nach der Dreiecksungleichung und unter Berücksichtigung des Hilfssatzes sowohl (6) als auch (7).

Gelte nun in (6) bzw. (7) für ein $z = z_0$ das Gleichheitszeichen. Da $z_0 = 0$ trivial ist, dürfen wir im folgenden $z_0 \neq 0$ voraussetzen. Die Gleichheitsaussage des benutzten Hilfssatzes und die bekannte Gleichheitsaussage bei der Dreiecksungleichung lassen dann die folgenden Feststellungen zu:

1) Der Fall $\mathfrak{a} = \mathfrak{b}$.

Nur die Behandlung von (6) ist interessant.

Notwendig und hinreichend für die Gültigkeit des Gleichheitszeichens in $z = z_0$ sind die Bedingungen

$$\mathfrak{f}_2(z_0) = -\mathfrak{g}_2(z_0), \quad |\mathfrak{f}_2(z_0)| = 1 \tag{14}$$

und

$$\mathfrak{a}^* \mathfrak{f}_2(z_0) = 0, \quad \text{falls } n > 1. \tag{15}$$

Wegen $|\mathfrak{f}_2(z_0)| = 1$ ergibt sich aus (9)

$$|\mathfrak{f}_1(z_0)| = 1.$$

Da $\mathfrak{f}_1(z)$ eine beschränkte Funktionenspalte ist, muß also $\mathfrak{f}_1(z)$ eine konstante Spalte vom Absolutbetrag 1 sein. Entsprechendes gilt für $\mathfrak{g}_1(z)$. Die Gesamtheit der das Gleichheitszeichen in z_0 erzielenden Funktionen $\mathfrak{f}, \mathfrak{g}$ läßt sich also so beschreiben: $\mathfrak{e}$ sei eine beliebige Spalte vom Absolutbetrag 1, die für $n > 1$

$$\mathfrak{a}^* \mathfrak{e} = 0$$

erfüllt $(\mathfrak{e} = \mathfrak{f}_2(z_0))$. Mit den Spalten

$$\mathfrak{f}_1 = \Gamma(\bar{z}_0 \mathfrak{a}) \frac{\mathfrak{e} - \bar{z}_0 \mathfrak{a}}{1 - \mathfrak{a}^* z_0 \mathfrak{e}}, \quad \mathfrak{g}_1 = -\Gamma(\bar{z}_0 \mathfrak{a}) \frac{\mathfrak{e} + \bar{z}_0 \mathfrak{a}}{1 + \mathfrak{a}^* z_0 \mathfrak{e}}$$

vom Absolutbetrag 1 sind

$$\mathfrak{f}(z) \equiv \Gamma(\mathfrak{a}) \frac{\mathfrak{f}_1 \cdot z + \mathfrak{a}}{1 + \mathfrak{a}^* \mathfrak{f}_1 \cdot z}, \quad \mathfrak{g}(z) \equiv \Gamma(\mathfrak{a}) \frac{\mathfrak{g}_1 \cdot z + \mathfrak{a}}{1 + \mathfrak{a}^* \mathfrak{g}_1 \cdot z}$$

zu bilden.

2) Der Fall $\mathfrak{a} \neq \mathfrak{b}$. Es ist dann auch*

$$\frac{\mathfrak{a}}{1 - |\mathfrak{a}|^2 |z_0|^2} \neq \frac{\mathfrak{b}}{1 - |\mathfrak{b}|^2 |z_0|^2}.$$

Aus

$$\frac{\mathfrak{a}}{1 - |\mathfrak{a}|^2 |z_0|^2} = \frac{\mathfrak{b}}{1 - |\mathfrak{b}|^2 |z_0|^2} \tag{16}$$

folgt nämlich

$$\frac{|\mathfrak{a}|}{1 - |\mathfrak{a}|^2 |z_0|^2} = \frac{|\mathfrak{b}|}{1 - |\mathfrak{b}|^2 |z_0|^2},$$

d. h.

$$(|\mathfrak{a}| - |\mathfrak{b}|)(1 + |\mathfrak{a}||\mathfrak{b}||z_0|^2) = 0,$$

d. h.

$$|\mathfrak{a}| = |\mathfrak{b}|,$$

und nun liefert (16)

$$\mathfrak{a} = \mathfrak{b}. \tag{17}$$

Die Gültigkeit des Gleichheitszeichens in (6) bzw. (7) für ein $z = z_0 \neq 0$ erfordert nun neben den unter 1) gefundenen Bedingungen zunächst einmal analog zu (15):

$$\mathfrak{b}^* \mathfrak{g}_2(z_0) = 0, \qquad \text{falls } n > 1, \tag{18}$$

außerdem aber, daß

$$z_0\left(\frac{1-|\mathfrak{a}|^2}{1-|\mathfrak{a}|^2|z_0|^2}\, \Gamma^{-1}(\mathfrak{a})\, \Gamma(\bar{z}_0 \mathfrak{a}) + \frac{1-|\mathfrak{b}|^2}{1-|\mathfrak{b}|^2|z_0|^2}\, \Gamma^{-1}(\mathfrak{b})\, \Gamma(\bar{z}_0 \mathfrak{b})\right) \mathfrak{f}_2(z_0)$$

ein positives bzw. negatives Vielfaches von

$$(1-|z_0|^2)\left(\frac{\mathfrak{a}}{1-|\mathfrak{a}|^2|z_0|^2} - \frac{\mathfrak{b}}{1-|\mathfrak{b}|^2|z_0|^2}\right)$$

ist (bei (7) setzen wir schon voraus, daß die rechte Seite in $z = z_0$ nicht negativ ist). Für $n = 1$ ist das offenbar zu erfüllen, $\mathfrak{f}_2(z_0)$ ist sogar eindeutig bestimmt. Für $n > 1$ aber ist das unerfüllbar, wie wir nun noch zeigen wollen. Nach Verwendung von (14), (15), (18) lautet die Bedingung nämlich:

$$z_0\left(M(z_0, \mathfrak{a}) + M(z_0, \mathfrak{b})\right) \mathfrak{f}_2(z_0) = \lambda \cdot (1-|z_0|^2)\left(\frac{\mathfrak{a}}{1-|\mathfrak{a}|^2|z_0|^2} - \frac{\mathfrak{b}}{1-|\mathfrak{b}|^2|z_0|^2}\right)$$

mit reellem $\lambda \neq 0$. Dies ist aber, wie man leicht nachprüft, ein Widerspruch zu

$$\mathfrak{a}^* \mathfrak{f}_2(z_0) = \mathfrak{b}^* \mathfrak{f}_2(z_0) = 0$$

(man beachte $\mathfrak{a} \neq \mathfrak{b}$).

Damit ist der Satz vollständig bewiesen. Es sind darüber hinaus sogar die Paare $\mathfrak{f}(z)$, $\mathfrak{g}(z)$ von Funktionenspalten, die in (6) bzw. (7) das Gleichheitszeichen bewirken, genau beschrieben worden.

Bemerkungen:

1. Wir haben gesehen, daß die Abschätzungen (6) und (7) keinesfalls scharf sind, wenn $z \neq 0$, $n > 1$ und $\mathfrak{a} \neq \mathfrak{b}$ sind. In speziellen Fällen können wir bessere Abschätzungen angeben. Für $\mathfrak{a} = -\mathfrak{b}$ z. B. ist

$$2\,\frac{|\mathfrak{a}|-|z|}{1-|\mathfrak{a}||z|} \leqq |\mathfrak{f}(z) - \mathfrak{g}(z)| \leqq 2\,\frac{|\mathfrak{a}|+|z|}{1+|\mathfrak{a}||z|}$$ [1])

besser, wie man leicht ausrechnet. Dies sind sogar die scharfen Schranken, da die entsprechenden Schranken für $n = 1$ scharf sind.

2. Die Abschätzungen (6) und (7) bleiben richtig, wenn $|\mathfrak{a}| = 1$ oder $|\mathfrak{b}| = 1$ ist. Es muß dabei

$$M(z, \mathfrak{a}) = 0 \quad \text{für} \quad |\mathfrak{a}| = 1$$

gesetzt werden.

3. Seien $\mathfrak{f}(z)$, $\mathfrak{g}(z)$ beschränkte Funktionenspalten mit $\mathfrak{f}(0) \neq \mathfrak{g}(0)$. Dann gibt es wegen der Stetigkeit eine Kreisscheibe $|z| < r$ um den Nullpunkt, in der $\mathfrak{f}(z)$ und $\mathfrak{g}(z)$ in ihren Werten nicht übereinstimmen können:

$$\mathfrak{f}(z) \neq \mathfrak{g}(z) \quad \text{in} \quad |z| < r.$$

Die Ungleichung (7) zeigt nun, daß bei vorgegebenen Spalten $\mathfrak{a} = \mathfrak{f}(0)$, $\mathfrak{b} = \mathfrak{g}(0)$ $(|\mathfrak{a}| < 1, |\mathfrak{b}| < 1, \mathfrak{a} \neq \mathfrak{b})$ eine gemeinsame solche Kreisscheibe existiert. Der

Radius der größtmöglichen derartigen Kreisscheibe werde $R(\mathfrak{a}, \mathfrak{b})$ genannt, also

$$R(\mathfrak{a}, \mathfrak{b}) = \max \left(r \,\middle|\, \begin{matrix} \mathfrak{f}(z) \neq \mathfrak{g}(z) \quad \text{in} \quad |z| < r \quad \text{für alle} \\ \text{beschränkten } \mathfrak{f}, \mathfrak{g} \quad \text{mit} \quad \mathfrak{f}(0) = \mathfrak{a},\ \mathfrak{g}(0) = \mathfrak{b} \end{matrix} \right).$$

Wenn die kleinste positive Zahl r, für die

$$(1 - r^2) \left| \frac{\mathfrak{a}}{1 - |\mathfrak{a}|^2 r^2} - \frac{\mathfrak{b}}{1 - |\mathfrak{b}|^2 r^2} \right| - r(M(r, \mathfrak{a}) + M(r, \mathfrak{b})) = 0$$

ist, $r(\mathfrak{a}, \mathfrak{b})$ genannt wird, so ist

$$R(\mathfrak{a}, \mathfrak{b}) \geqq r(\mathfrak{a}, \mathfrak{b}),$$

und für $n = 1$ gilt sogar das Gleichheitszeichen, wie Satz 1 lehrt. — Auf einem ganz anderen Wege läßt sich nun aber die Zahl $R(\mathfrak{a}, \mathfrak{b})$ auch für $n > 1$ ermitteln. Es gilt nämlich der noch weiter gehende

Satz 2. *Seien* $\mathfrak{a}, \mathfrak{b}$ *gegebene Spalten mit* $|\mathfrak{a}| < 1$, $|\mathfrak{b}| < 1$, $\mathfrak{a} \neq \mathfrak{b}$. *Es ist*

$$R(\mathfrak{a}, \mathfrak{b}) = \left(\frac{|1 - \mathfrak{b}^* \mathfrak{a}| - v(\mathfrak{a})\, v(\mathfrak{b})}{|1 - \mathfrak{b}^* \mathfrak{a}| + v(\mathfrak{a})\, v(\mathfrak{b})} \right)^{\frac{1}{2}}. \tag{19}$$

Für beliebige beschränkte Funktionenspalten $\mathfrak{f}(z)$, $\mathfrak{g}(z)$ *mit* $\mathfrak{f}(0) = \mathfrak{a}$, $\mathfrak{g}(0) = \mathfrak{b}$ *gilt sogar:*

$$\mathfrak{f}(z_1) \neq \mathfrak{g}(z_2) \quad \textit{für} \quad |z_1| < R(\mathfrak{a}, \mathfrak{b}),\ |z_2| < R(\mathfrak{a}, \mathfrak{b}).$$

Beweis. Die durch das Hermitesche Linienelement

$$\frac{1}{1 - |\mathfrak{z}|^2} |\Gamma(\mathfrak{z})\, d\mathfrak{z}|$$

im Innern der Einheitshyperkugel $|\mathfrak{z}| < 1$ verwirklichte nicht-euklidische Geometrie hat als Bewegungsgruppe die Gruppe der analytischen Automorphismen der Hyperkugel. $I(\mathfrak{z}_1, \mathfrak{z}_2) = \left| \Gamma(\mathfrak{z}_2) \frac{\mathfrak{z}_1 - \mathfrak{z}_2}{1 - \mathfrak{z}_2^* \mathfrak{z}_1} \right|$ ist ein invariantes Maß für die die beiden Punkte $\mathfrak{z}_1, \mathfrak{z}_2$ verbindende Geodätische. Es handelt sich allerdings nicht um die Länge $L(\mathfrak{z}_1, \mathfrak{z}_2)$ der Geodätischen. L ist vielmehr durch eine gewisse monotone Funktion dieser Invarianten I gegeben, nämlich

$$L = \operatorname{Artgh} I.$$

Wir brauchen aber nur Feststellungen über Längengleichheit zu machen, und darum können wir obige Invariante I benutzen. Schließlich ist zu bemerken, daß die durch den Nullpunkt gehenden Geraden, soweit sie in der Einheitshyperkugel verlaufen, Geodätische sind.

Mit diesen Kenntnissen ist es leicht, den Beweis der Existenz und Eindeutigkeit einer Spalte $\mathfrak{c} = \mathfrak{c}(\mathfrak{a}, \mathfrak{b})$, die

$$\Gamma(\mathfrak{c}) \frac{\mathfrak{a} - \mathfrak{c}}{1 - \mathfrak{c}^* \mathfrak{a}} = -\Gamma(\mathfrak{c}) \frac{\mathfrak{b} - \mathfrak{c}}{1 - \mathfrak{c}^* \mathfrak{b}} \tag{20}$$

erfüllt, zu führen.

Existenzbeweis. Der nicht-euklidische Mittelpunkt der geodätischen Verbindung von $\mathfrak{a}$ und $\mathfrak{b}$ sei $\mathfrak{c}$ genannt. Die nicht-euklidische Bewegung

$$\mathfrak{z} \to \Gamma(\mathfrak{c}) \frac{\mathfrak{z} - \mathfrak{c}}{1 - \mathfrak{c}^* \mathfrak{z}} \tag{21}$$

überführt die von $\mathfrak{a}$ nach $\mathfrak{b}$ führende Geodätische in ein Geradenstück durch den Nullpunkt, dessen Endpunkte

$$\mathfrak{a}_1 = \Gamma(\mathfrak{c}) \frac{\mathfrak{a} - \mathfrak{c}}{1 - \mathfrak{c}^* \mathfrak{a}}, \qquad \mathfrak{b}_1 = \Gamma(\mathfrak{c}) \frac{\mathfrak{b} - \mathfrak{c}}{1 - \mathfrak{c}^* \mathfrak{b}} \tag{22}$$

vom Nullpunkt die gleiche nicht-euklidische Entfernung haben müssen: $|\mathfrak{a}_1| = |\mathfrak{b}_1|$. Dann muß aber

$$\mathfrak{a}_1 = -\mathfrak{b}_1 \tag{23}$$

sein.

Eindeutigkeitsbeweis. Sei $\mathfrak{c}$ irgendein Punkt der Einheitshyperkugel, für den (20) gilt. Die nicht-euklidische Bewegung (21) überführt ihn in den Nullpunkt, während $\mathfrak{a}$ bzw. $\mathfrak{b}$ in $\mathfrak{a}_1$ bzw. $\mathfrak{b}_1$ (s. (22)) übergehen. Für die Bilder der Punkte $\mathfrak{c}, \mathfrak{a}, \mathfrak{b}$, nämlich $\mathfrak{o}, \mathfrak{a}_1, \mathfrak{b}_1$ gilt wegen (20) offenbar: $\mathfrak{o}$ ist nicht-euklidischer Mittelpunkt der Geodätischen von $\mathfrak{a}_1$ nach $\mathfrak{b}_1$. Also gilt das Analoge für die Urbilder.

Ohne die Kenntnisse aus jener nicht-euklidischen Geometrie läßt sich der Beweis auch leicht durch direkte Rechnung führen. Man erhält dabei einen expliziten Ausdruck für $\mathfrak{c}(\mathfrak{a}, \mathfrak{b})$. Aus (20) folgt insbesondere

$$1 - \left|\Gamma(\mathfrak{c}) \frac{\mathfrak{a} - \mathfrak{c}}{1 - \mathfrak{c}^* \mathfrak{a}}\right|^2 = 1 - \left|\Gamma(\mathfrak{c}) \frac{\mathfrak{b} - \mathfrak{c}}{1 - \mathfrak{c}^* \mathfrak{b}}\right|^2,$$

d. h.

$$\frac{(1 - |\mathfrak{c}|^2)(1 - |\mathfrak{a}|^2)}{|1 - \mathfrak{c}^* \mathfrak{a}|^2} = \frac{(1 - |\mathfrak{c}|^2)(1 - |\mathfrak{b}|^2)}{|1 - \mathfrak{c}^* \mathfrak{b}|^2},$$

d. h.

$$v(\mathfrak{b})(1 - \mathfrak{c}^* \mathfrak{a}) = \varepsilon\, v(\mathfrak{a})(1 - \mathfrak{c}^* \mathfrak{b}), \qquad |\varepsilon| = 1. \tag{24}$$

Dies wird in (20) verwandt:

$$v(\mathfrak{b})(\mathfrak{a} - \mathfrak{c}) + \varepsilon\, v(\mathfrak{a})(\mathfrak{b} - \mathfrak{c}) = 0, \tag{25}$$

d. h.

$$\big(v(\mathfrak{b}) + \varepsilon\, v(\mathfrak{a})\big)\mathfrak{c} = v(\mathfrak{b})\mathfrak{a} + \varepsilon\, v(\mathfrak{a})\mathfrak{b}. \tag{26}$$

(20) ist gleichbedeutend mit (24), (26). Zur Berechnung von ε wird nun (26) in die leicht umgeformte Gl. (24) eingesetzt:

$$\begin{cases} v(\mathfrak{b})\big(v(\mathfrak{b}) + \varepsilon\, v(\mathfrak{a})\big) - v(\mathfrak{b})\big(v(\mathfrak{b})\,\mathfrak{a}^*\mathfrak{a} + \varepsilon\, v(\mathfrak{a})\,\mathfrak{a}^*\mathfrak{b}\big) \\ \qquad = \bar{\varepsilon}\, v(\mathfrak{a})\big(v(\mathfrak{b}) + \varepsilon\, v(\mathfrak{a})\big) - \bar{\varepsilon}\, v(\mathfrak{a})\big(v(\mathfrak{b})\,\mathfrak{b}^*\mathfrak{a} + \varepsilon\, v(\mathfrak{a})\,\mathfrak{b}^*\mathfrak{b}\big), \end{cases} \tag{27}$$

d. h.

$$\varepsilon(1 - \mathfrak{a}^*\mathfrak{b}) = \bar{\varepsilon}(1 - \mathfrak{b}^*\mathfrak{a}).$$

Es ist also $\varepsilon(1 - \mathfrak{a}^*\mathfrak{b})$ reell, und da $|\varepsilon| = 1$ sein soll:

$$\varepsilon = \pm \frac{1 - \mathfrak{b}^*\mathfrak{a}}{|1 - \mathfrak{b}^*\mathfrak{a}|}. \tag{28}$$

Wir müssen jetzt noch in (26) $|c|<1$ garantieren, d.h.

$$|v(b)+\varepsilon v(a)|>|v(b)\,a+\varepsilon v(a)\,b|. \tag{29}$$

Wegen (29) ist $v(b)+\varepsilon v(a)\neq 0$ mitgarantiert, so daß aus (27) nach Division durch $v(b)+\varepsilon v(a)$ wieder (24) folgt. (29) besagt:

$$v^2(b)+v^2(a)+2v(a)\,v(b)\,\Re\varepsilon>v^2(b)\,|a|^2+v^2(a)\,|b|^2+2v(a)\,v(b)\,\Re(a^*\varepsilon\,b),$$

d.h.

$$v(a)\,v(b)+\Re\left((1-a^*b)\,\varepsilon\right)>0$$

und unter Verwendung von (28):

$$v(a)\,v(b)\pm|1-b^*a|>0.$$

Von den beiden Vorzeichen ist das untere unbrauchbar, da bekanntlich

$$\left|\Gamma(b)\,\frac{a-b}{1-b^*a}\right|^2=1-\left(\frac{v(a)\,v(b)}{|1-b^*a|}\right)^2>0$$

ist. Wir haben also zusammenfassend das Resultat:

$$c=c(a,b)=\frac{v(b)\,a+\varepsilon v(a)\,b}{v(b)+\varepsilon v(a)},\qquad \varepsilon=\frac{1-b^*a}{|1-b^*a|}. \tag{30}$$

Werde $\left(c=c(a,b)\right)$

$$\left|\Gamma(c)\,\frac{a-c}{1-c^*a}\right|=\varrho(a,b)\left(=\left|\Gamma(c)\,\frac{b-c}{1-c^*b}\right|\right)$$

gesetzt. Die beiden nicht-euklidischen Hypervollkugeln

$$\left|\Gamma(a)\,\frac{\mathfrak{z}-a}{1-a^*\mathfrak{z}}\right|\leqq\varrho,\qquad \left|\Gamma(b)\,\frac{\mathfrak{z}-b}{1-b^*\mathfrak{z}}\right|\leqq\varrho \tag{31}$$

haben, wie geometrisch klar ist, bei obiger Wahl von $\varrho=\varrho(a,b)$ nur einen einzigen Punkt gemeinsam, und zwar den Punkt $\mathfrak{z}=c$. Will man auch diese Aussage nochmal explizit rechnerisch nachprüfen, so tut man dies am besten nach Transformation mittels des Hyperkugelautomorphismus (21), der a, b in a_1, $-a_1$ überführt (s. (22) und (23)). Wenn die Bilder der Punkte $\mathfrak{z}$ mit $\mathfrak{w}$ bezeichnet werden, so lauten die Ungleichungen (31) wegen bekannter Invarianzeigenschaften ihrer linken Seiten:

$$\left|\Gamma(a_1)\,\frac{\mathfrak{w}-a_1}{1-a_1^*\mathfrak{w}}\right|\leqq|a_1|,\qquad \left|\Gamma(a_1)\,\frac{\mathfrak{w}+a_1}{1+a_1^*\mathfrak{w}}\right|\leqq|a_1|. \tag{32}$$

Man beachte nämlich, daß $\varrho=|a_1|$ ist. Hieraus muß $\mathfrak{w}=0$ gefolgert werden. (32) lautet:

$$\frac{1-|\mathfrak{w}|^2}{|1-a_1^*\mathfrak{w}|^2}\geqq 1,\qquad \frac{1-|\mathfrak{w}|^2}{|1+a_1^*\mathfrak{w}|^2}\geqq 1,$$

d.h.

$$|\mathfrak{w}|^2-2\Re(a_1^*\mathfrak{w})+|a_1^*\mathfrak{w}|^2\leqq 0,\qquad |\mathfrak{w}|^2+2\Re(a_1^*\mathfrak{w})+|a_1^*\mathfrak{w}|^2\leqq 0,$$

d.h.

$$|\mathfrak{w}|^2+2|\Re(a_1^*\mathfrak{w})|+|a_1^*\mathfrak{w}|^2\leqq 0,$$

woraus unmittelbar $\mathfrak{w}=0$ folgt.

Aus (4) ergibt sich

$$\left|\Gamma(\mathfrak{a})\,\frac{\mathfrak{f}(z_1)-\mathfrak{a}}{1-\mathfrak{a}^*\mathfrak{f}(z_1)}\right|<\varrho,\qquad \left|\Gamma(\mathfrak{b})\,\frac{\mathfrak{g}(z_2)-\mathfrak{b}}{1-\mathfrak{b}^*\mathfrak{g}(z_2)}\right|<\varrho$$

für alle z_1, z_2 mit $|z|<\varrho$, $|z_2|<\varrho$. Es ist also nach obigen Feststellungen

$$\mathfrak{f}(z_1)\neq\mathfrak{g}(z_2)\qquad\text{für}\qquad |z_1|<\varrho,\ |z_2|<\varrho. \tag{33}$$

Sei nun für ein z_0 auf $|z_0|=\varrho$

$$\mathfrak{f}(z_0)=\mathfrak{g}(z_0).$$

Es ist

$$\left|\Gamma(\mathfrak{a})\,\frac{\mathfrak{f}(z_0)-\mathfrak{a}}{1-\mathfrak{a}^*\mathfrak{f}(z_0)}\right|\leqq\varrho,\qquad \left|\Gamma(\mathfrak{b})\,\frac{\mathfrak{g}(z_0)-\mathfrak{b}}{1-\mathfrak{b}^*\mathfrak{g}(z_0)}\right|\leqq\varrho. \tag{34}$$

Nach obigen Feststellungen muß also

$$\mathfrak{f}(z_0)=\mathfrak{g}(z_0)=\mathfrak{c} \tag{35}$$

sein. In den Ungleichungen (34) gelten demnach die Gleichheitszeichen, was wiederum bekanntlich

$$\Gamma(\mathfrak{a})\,\frac{\mathfrak{f}(z)-\mathfrak{a}}{1-\mathfrak{a}^*\mathfrak{f}(z)}\equiv\mathfrak{d}z,\qquad \Gamma(\mathfrak{b})\,\frac{\mathfrak{g}(z)-\mathfrak{b}}{1-\mathfrak{b}^*\mathfrak{g}(z)}\equiv\mathfrak{e}z\qquad(|\mathfrak{d}|=|\mathfrak{e}|=1) \tag{36}$$

erfordert. Die Verwendung von (35) in (36) schließlich bestimmt $\mathfrak{d}$ und $\mathfrak{e}$ eindeutig:

$$\Gamma(\mathfrak{a})\,\frac{\mathfrak{c}-\mathfrak{a}}{1-\mathfrak{a}^*\mathfrak{c}}=\mathfrak{d}z_0,\qquad \Gamma(\mathfrak{b})\,\frac{\mathfrak{c}-\mathfrak{b}}{1-\mathfrak{b}^*\mathfrak{c}}=\mathfrak{e}z_0.$$

Daß diese Spalten $\mathfrak{d}, \mathfrak{e}$ tatsächlich den absoluten Betrag 1 haben, ist klar, denn es ist

$$\left|\Gamma(\mathfrak{a})\,\frac{\mathfrak{c}-\mathfrak{a}}{1-\mathfrak{a}^*\mathfrak{c}}\right|=\left|\Gamma(\mathfrak{c})\,\frac{\mathfrak{a}-\mathfrak{c}}{1-\mathfrak{c}^*\mathfrak{a}}\right|=|z_0|,\qquad \left|\Gamma(\mathfrak{b})\,\frac{\mathfrak{c}-\mathfrak{b}}{1-\mathfrak{b}^*\mathfrak{c}}\right|=\left|\Gamma(\mathfrak{c})\,\frac{\mathfrak{b}-\mathfrak{c}}{1-\mathfrak{c}^*\mathfrak{b}}\right|=|z_0|.$$

Es ist also

$$\varrho(\mathfrak{a},\mathfrak{b})=R(\mathfrak{a},\mathfrak{b}).$$

Was allein jetzt nachzuweisen bleibt, um den Beweis des Satzes zu vollenden, ist die Formel

$$\left|\Gamma(\mathfrak{c})\,\frac{\mathfrak{a}-\mathfrak{c}}{1-\mathfrak{c}^*\mathfrak{a}}\right|=\left(\frac{|1-\mathfrak{b}^*\mathfrak{a}|-v(\mathfrak{a})\,v(\mathfrak{b})}{|1-\mathfrak{b}^*\mathfrak{a}|+v(\mathfrak{a})\,v(\mathfrak{b})}\right)^{\frac{1}{2}}. \tag{37}$$

Es ist

$$\begin{aligned}\frac{1-|\mathfrak{c}|^2}{|1-\mathfrak{a}^*\mathfrak{c}|^2}&=\frac{|v(\mathfrak{b})+\varepsilon v(\mathfrak{a})|^2-|v(\mathfrak{b})\,\mathfrak{a}+\varepsilon v(\mathfrak{a})\,\mathfrak{b}|^2}{|v(\mathfrak{b})+\varepsilon v(\mathfrak{a})-v(\mathfrak{b})\,|\mathfrak{a}|^2-\varepsilon v(\mathfrak{a})\,\mathfrak{a}^*\mathfrak{b}|^2}\\&=\frac{2v^2(\mathfrak{a})\,v^2(\mathfrak{b})+2v(\mathfrak{a})\,v(\mathfrak{b})\,\Re((1-\mathfrak{a}^*\mathfrak{b})\,\varepsilon)}{|v^2(\mathfrak{a})\,v(\mathfrak{b})+v(\mathfrak{a})\,(1-\mathfrak{a}^*\mathfrak{b})\,\varepsilon|^2}\\&=\frac{2v^2(\mathfrak{a})\,v^2(\mathfrak{b})+2v(\mathfrak{a})\,v(\mathfrak{b})|\,1-\mathfrak{b}^*\mathfrak{a}|}{(v^2(\mathfrak{a})\,v(\mathfrak{b})+v(\mathfrak{a})|\,1-\mathfrak{b}^*\mathfrak{a}|)^2}\\&=\frac{1}{v^2(\mathfrak{a})}\cdot\frac{2v(\mathfrak{a})\,v(\mathfrak{b})}{v(\mathfrak{a})\,v(\mathfrak{b})+|\,1-\mathfrak{b}^*\mathfrak{a}|},\end{aligned}$$

woraus sich (37) wegen

$$\left|\Gamma(\mathfrak{c})\,\frac{\mathfrak{a}-\mathfrak{c}}{1-\mathfrak{c}^*\mathfrak{a}}\right|^2=1-v^2(\mathfrak{a})\,\frac{1-|\mathfrak{c}|^2}{|1-\mathfrak{a}^*\mathfrak{c}|^2}$$

unmittelbar ergibt. — Man hätte (37) natürlich auch aus der geometrischen Kenntnis

$$\text{(38)} \qquad \operatorname{Artgh}\left|\Gamma(\mathfrak{c})\,\frac{\mathfrak{a}-\mathfrak{c}}{1-\mathfrak{c}^*\mathfrak{a}}\right| = \frac{1}{2}\operatorname{Artgh}\left|\Gamma(\mathfrak{b})\,\frac{\mathfrak{a}-\mathfrak{b}}{1-\mathfrak{b}^*\mathfrak{a}}\right|$$

heraus vermittels der bekannten Funktionalgleichung, die zwischen $\operatorname{tgh} x$ und $\operatorname{tgh}\frac{x}{2}$ besteht, gewinnen können.

Satz 3. *Seien* $\mathfrak{a}, \mathfrak{b}$ *gegebene Spalten mit* $|\mathfrak{a}|<1$, $|\mathfrak{b}|<1$. *Für beliebige beschränkte Funktionenspalten* $\mathfrak{f}(z)$, $\mathfrak{g}(z)$ *mit* $\mathfrak{f}(0)=\mathfrak{a}$, $\mathfrak{g}(0)=\mathfrak{b}$ *gilt*

$$\text{(39)} \qquad \left|\Gamma(\mathfrak{g}(z))\,\frac{\mathfrak{f}(z)-\mathfrak{g}(z)}{1-\mathfrak{g}^*(z)\,\mathfrak{f}(z)}\right| \leqq 2\cdot\frac{\dfrac{R+|z|}{1+R|z|}}{1+\left(\dfrac{R+|z|}{1+R|z|}\right)^2} \qquad (R=R(\mathfrak{a},\mathfrak{b})).$$

Die Schranke ist für jedes z des Einheitskreises scharf.

Beweis. Mit der Spalte $\mathfrak{c}=\mathfrak{c}(\mathfrak{a},\mathfrak{b})$ aus (30), die hier auch für $\mathfrak{a}=\mathfrak{b}$ Verwendung findet $(\mathfrak{c}(\mathfrak{a},\mathfrak{a})=\mathfrak{a})$, wird

$$\Gamma(\mathfrak{c})\,\frac{\mathfrak{f}(z)-\mathfrak{c}}{1-\mathfrak{c}^*\mathfrak{f}(z)}=\mathfrak{f}_1(z), \qquad \Gamma(\mathfrak{c})\,\frac{\mathfrak{g}(z)-\mathfrak{c}}{1-\mathfrak{c}^*\mathfrak{g}(z)}=\mathfrak{g}_1(z)$$

gesetzt. Wie in (22) sei

$$\Gamma(\mathfrak{c})\,\frac{\mathfrak{a}-\mathfrak{c}}{1-\mathfrak{c}^*\mathfrak{a}}=\mathfrak{a}_1\left(=-\,\Gamma(\mathfrak{c})\,\frac{\mathfrak{b}-\mathfrak{c}}{1-\mathfrak{c}^*\mathfrak{b}}\right);$$

dann ist (man beachte (20))

$$\mathfrak{f}_1(0)=-\,\mathfrak{g}_1(0)=\mathfrak{a}_1.$$

In

$$\mathfrak{f}_1(z)=\Gamma(\mathfrak{a}_1)\,\frac{z\,\mathfrak{f}_2(z)+\mathfrak{a}_1}{1+\mathfrak{a}_1^*\,z\,\mathfrak{f}_2(z)}, \qquad \mathfrak{g}_1(z)=\Gamma(\mathfrak{a}_1)\,\frac{z\,\mathfrak{g}_2(z)-\mathfrak{a}_1}{1-\mathfrak{a}_1^*\,z\,\mathfrak{g}_2(z)}$$

sind $\mathfrak{f}_2$, $\mathfrak{g}_2$ beschränkte Funktionenspalten, und es ist

$$\begin{aligned}
1-\left|\Gamma(\mathfrak{g}_1)\,\frac{\mathfrak{f}_1-\mathfrak{g}_1}{1-\mathfrak{g}_1\mathfrak{f}_1}\right|^2 &= \frac{(1-|\mathfrak{g}_1|^2)(1-|\mathfrak{f}_1|^2)}{|1-\mathfrak{g}_1^*\mathfrak{f}_1|^2}\\
&= \frac{(1-|\mathfrak{a}_1|^2)^2(1-|z|^2|\mathfrak{g}_2|^2)(1-|z|^2|\mathfrak{f}_2|^2)}{|(1-\bar z\,\mathfrak{g}_2^*\mathfrak{a}_1)(1+\mathfrak{a}_1^*z\mathfrak{f}_2)-(\bar z\,\mathfrak{g}_2^*-\mathfrak{a}_1^*)\Gamma^2(\mathfrak{a}_1)(z\mathfrak{f}_2+\mathfrak{a}_1)|^2}\\
&= \frac{(1-|\mathfrak{a}_1|^2)^2(1-|z|^2|\mathfrak{g}_2|^2)(1-|z|^2|\mathfrak{f}_2|^2)}{|1+|\mathfrak{a}_1|^2-2\bar z\,\mathfrak{g}_2^*\mathfrak{a}_1+2\mathfrak{a}_1^*z\mathfrak{f}_2-2z\bar z\,\mathfrak{g}_2^*\mathfrak{a}_1\mathfrak{a}_1^*\mathfrak{f}_2-z\bar z\,v^2(\mathfrak{a}_1)\,\mathfrak{g}_2^*\mathfrak{f}_2|^2}\\
&\geqq \frac{(1-|\mathfrak{a}_1|^2)^2(1-|z|^2|\mathfrak{g}_2|^2)(1-|z|^2|\mathfrak{f}_2|^2)}{(1+|\mathfrak{a}_1|^2+2|z|(|\mathfrak{g}_2^*\mathfrak{a}_1|+|\mathfrak{a}_1^*\mathfrak{f}_2|)+|z|^2(2|\mathfrak{g}_2^*\mathfrak{a}_1||\mathfrak{a}_1^*\mathfrak{f}_2|+v^2(\mathfrak{a}_1)|\mathfrak{g}_2^*\mathfrak{f}_2|))^2}.
\end{aligned}$$

Da $|\mathfrak{g}_2^*\mathfrak{a}_1|\leqq|\mathfrak{a}_1||\mathfrak{g}_2|\leqq|\mathfrak{a}_1|$, $|\mathfrak{a}_1^*\mathfrak{f}_2|\leqq|\mathfrak{a}_1||\mathfrak{f}_2|\leqq|\mathfrak{a}_1|$ ist, folgt:

$$\begin{aligned}
1-\left|\Gamma(\mathfrak{g}_1)\,\frac{\mathfrak{f}_1-\mathfrak{g}_1}{1-\mathfrak{g}_1^*\mathfrak{f}_1}\right|^2 &\geqq \left(\frac{(1-|\mathfrak{a}_1|^2)(1-|z|^2)}{1+|\mathfrak{a}_1|^2+4|\mathfrak{a}_1||z|+(1+|\mathfrak{a}_1|^2)|z|^2}\right)^2\\
&= \left(\frac{1-\left(\dfrac{|\mathfrak{a}_1|+|z|}{1+|\mathfrak{a}_1||z|}\right)^2}{1+\left(\dfrac{|\mathfrak{a}_1|+|z|}{1+|\mathfrak{a}_1||z|}\right)^2}\right)^2
\end{aligned}$$

$$\text{(40)} \qquad \left|\Gamma(\mathfrak{g}_1)\,\frac{\mathfrak{f}_1-\mathfrak{g}_1}{1-\mathfrak{g}_1^*\mathfrak{f}_1}\right| \leqq 2\cdot\frac{\dfrac{|\mathfrak{a}_1|+|z|}{1+|\mathfrak{a}_1||z|}}{1+\left(\dfrac{|\mathfrak{a}_1|+|z|}{1+|\mathfrak{a}_1||z|}\right)^2}.$$

Die bekannte Invarianzeigenschaft

$$\left|\Gamma(\mathfrak{g}_1)\,\frac{\mathfrak{f}_1-\mathfrak{g}_1}{1-\mathfrak{g}_1^*\mathfrak{f}_1}\right| = \left|\Gamma(\mathfrak{g})\,\frac{\mathfrak{f}-\mathfrak{g}}{1-\mathfrak{g}^*\mathfrak{f}}\right|$$

führt jetzt unmittelbar zu (39), wenn man

$$|\mathfrak{a}_1| = R(\mathfrak{a},\mathfrak{b})$$

beachtet.

Das Gleichheitszeichen in (39), d.h. in (40), steht, wie man leicht einsieht, für ein $z=z_0\neq 0$ genau dann, wenn gilt:

$$\mathfrak{f}_2(z)\equiv \mathfrak{e}\,(|\mathfrak{e}|=1),\qquad \mathfrak{g}_2(z)=-\mathfrak{f}_2(z),$$

und, falls $\mathfrak{a}_1\neq 0$,

$$\mathfrak{e}=\frac{\bar{z}_0\mathfrak{a}_1}{|\bar{z}_0\mathfrak{a}_1|}.$$

Es besteht natürlich für $\left|\Gamma(\mathfrak{g}(z))\,\frac{\mathfrak{f}(z)-\mathfrak{g}(z)}{1-\mathfrak{g}^*(z)\,\mathfrak{f}(z)}\right|$ auch eine scharfe Abschätzung nach unten. Es wird vermutet, daß diese

$$\left|\Gamma(\mathfrak{g}(z))\,\frac{\mathfrak{f}(z)-\mathfrak{g}(z)}{1-\mathfrak{g}^*(z)\,\mathfrak{f}(z)}\right| \geqq 2\cdot\frac{\dfrac{R-|z|}{1-R|z|}}{1+\left(\dfrac{R-|z|}{1-R|z|}\right)^2}\qquad (R=R(\mathfrak{a},\mathfrak{b}))$$

lautet, solange $|z|\leqq R$ ist.

Satz 3 enthält als Spezialfall, nämlich für $\mathfrak{b}=-\mathfrak{a}$, $\mathfrak{g}(z)=-\mathfrak{f}(z)$, die bekannte Verallgemeinerung des Schwarzschen Lemmas[1])

$$|\mathfrak{f}(z)|\leqq\frac{|\mathfrak{a}|+|z|}{1+|\mathfrak{a}|\,|z|},\tag{41}$$

für $\mathfrak{a}=0$ das Schwarzsche Lemma selbst. Das Schwarzsche Lemma läßt sich natürlich in verschiedenster Weise verallgemeinern. Eine Möglichkeit, die einen besonders überraschenden Unterschied zwischen $n=1$ und $n>1$ erkennen läßt, zeigt der folgende Satz auf.

Satz 4. *Sei $\mathfrak{a}$ eine vorgegebene Spalte mit $|\mathfrak{a}|\leqq 1$. Für jede beschränkte Funktionenspalte $\mathfrak{f}(z)$ mit $\mathfrak{f}(0)=\mathfrak{a}$ gilt*

$$|\mathfrak{f}(z)-\mathfrak{a}|\leqq\begin{cases}(1-|\mathfrak{a}|^2)\,\dfrac{|z|}{1-|\mathfrak{a}|\,|z|} & \text{für } n=1\\ \text{und für } n>1,\ |z|\geqq|\mathfrak{a}|,\\ \sqrt{1-|\mathfrak{a}|^2}\,\dfrac{|z|}{\sqrt{1-|z|^2}} & \text{für } n>1,\ |z|\leqq|\mathfrak{a}|.\end{cases}\tag{42}$$

Die Abschätzung ist stets scharf.

Beweis. Der Fall $|\mathfrak{a}|=1$ ist trivial, und es kann also nun $|\mathfrak{a}|<1$ angenommen werden. Werde wieder

$$\mathfrak{f}(z)=\Gamma(\mathfrak{a})\,\frac{z\,\mathfrak{f}_1(z)+\mathfrak{a}}{1+\mathfrak{a}^*z\,\mathfrak{f}_1(z)}$$

gesetzt. Es ist

$$\begin{aligned}\mathfrak{f}-\mathfrak{a} &= \Gamma(\mathfrak{a})\left(\frac{z\mathfrak{f}_1+\mathfrak{a}}{1+\mathfrak{a}^* z\mathfrak{f}_1}-\mathfrak{a}\right)\\ &= \Gamma(\mathfrak{a})\,(\mathsf{E}-\mathfrak{a}\mathfrak{a}^*)\,\frac{z\mathfrak{f}_1}{1+\mathfrak{a}^* z\mathfrak{f}_1}\\ &= v^2(\mathfrak{a})\,\Gamma^{-1}(\mathfrak{a})\,\frac{z\mathfrak{f}_1}{1+\mathfrak{a}^* z\mathfrak{f}_1}\,.\end{aligned}$$

Für $n=1$ liegt die Abschätzung nun auf der Hand. Man sieht auch sofort ein, daß das Gleichheitszeichen für ein $z=z_0\neq 0$ genau dann erreicht wird wenn $\mathfrak{f}_1(z)\equiv\mathfrak{e}$ $(|\mathfrak{e}|=1)$ und, falls $\mathfrak{a}\neq 0$, $\mathfrak{e}=-\frac{\bar z_0\,\mathfrak{a}}{|\bar z_0\,\mathfrak{a}|}$ ist. Sei nun $n>1$.

$$\begin{aligned}|\mathfrak{f}-\mathfrak{a}|^2 &= \frac{v^2(\mathfrak{a})}{|1+\mathfrak{a}^* z\mathfrak{f}_1|^2}\,|z|^2\mathfrak{f}_1^*(\mathsf{E}-\mathfrak{a}\mathfrak{a}^*)\mathfrak{f}_1\\ &= v^2(\mathfrak{a})\,|z|^2\,\frac{|\mathfrak{f}_1|^2-|\mathfrak{a}^*\mathfrak{f}_1|^2}{|1+\mathfrak{a}^* z\mathfrak{f}_1|^2}\\ &\leqq v^2(\mathfrak{a})\,|z|^2\,\frac{|\mathfrak{f}_1|^2-|\mathfrak{a}^*\mathfrak{f}_1|^2}{(1-|z|\,|\mathfrak{a}^*\mathfrak{f}_1|)^2}\,.\end{aligned}$$

Es muß das Maximum der rechten Seite für beliebig in $|\mathfrak{f}_1|\leqq 1$ variierendes $\mathfrak{f}_1$ aufgesucht werden. Wir setzen

$$\mathfrak{f}_1=|\mathfrak{f}_1|\,\mathfrak{e},\qquad |\mathfrak{e}|=1,$$

und erhalten

$$\begin{aligned}|\mathfrak{f}-\mathfrak{a}|^2 &\leqq v^2(\mathfrak{a})\,|z|^2\,\frac{|\mathfrak{f}_1|^2(1-|\mathfrak{a}^*\mathfrak{e}|^2)}{(1-|z|\,|\mathfrak{f}_1|\,|\mathfrak{a}^*\mathfrak{e}|)^2}\\ &\leqq v^2(\mathfrak{a})\,|z|^2\,\frac{1-|\mathfrak{a}^*\mathfrak{e}|^2}{(1-|z|\,|\mathfrak{a}^*\mathfrak{e}|)^2}\,.\end{aligned}$$

Man prüft leicht nach, daß das Maximum der Funktion $\frac{1-x^2}{(1-|z|\,x)^2}$ $(0\leqq x\leqq|\mathfrak{a}|)$ für $x=|z|$ oder $x=|\mathfrak{a}|$ erreicht wird, je nachdem, ob $|z|\leqq|\mathfrak{a}|$ oder $|z|\geqq|\mathfrak{a}|$ ist. Damit ist (42) nachgewiesen. Über die Gültigkeit des Gleichheitszeichens gilt wörtlich das im Falle $n=1$ Gesagte, solange $|z_0|\geqq|\mathfrak{a}|$ ist. Sei nun aber $|z_0|<|\mathfrak{a}|$. Neben $|\mathfrak{f}_1|=1$, d.h. $\mathfrak{f}_1(z)\equiv\mathfrak{e}$ $(|\mathfrak{e}|=1)$, wird jetzt für die Gültigkeit des Gleichheitszeichens

$$\mathfrak{a}^*\mathfrak{e}=-\bar z_0$$

als notwendig und hinreichend erkannt, d.h.

$$\mathfrak{e}=-\frac{\bar z_0}{|\mathfrak{a}|^2}\,\mathfrak{a}+\mathfrak{b}$$

mit beliebiger Spalte $\mathfrak{b}$, die

$$\mathfrak{a}^*\mathfrak{b}=0,\qquad |\mathfrak{b}|^2=1-\frac{|z_0|^2}{|\mathfrak{a}|^2}$$

erfüllt.

Bonn, Mathematisches Institut der Universität

(Eingegangen am 2. März 1959)

HENRIKSEN, M.
Math. Zeitschr. 71, 427–435 (1959)

Multiplicative summability methods and the Stone-Čech compactification

By
MELVIN HENRIKSEN[1])

Introduction

A regular summability method φ (for real sequences) is called multiplicative *(b-multiplicative)* if whenever f, g are two (bounded) sequences summed by φ, then $\varphi(fg) = \varphi(f)\,\varphi(g)$. It is known [11], [17], p. 71 that the regular multiplicative matrix summability methods are submethods of the method φ_I corresponding to the identity matrix I (i.e., methods obtained from I by deleting infinitely many of its rows). It was conjectured in [1] that every regular b-multiplicative matrix summability method is equivalent for bounded sequences to a submethod of φ_I. Below (3.3), we disprove this conjecture by exhibiting a countable decreasing chain of submethods of φ_I such that the intersection of their bounded summability fields is not the bounded summability field of a submethod of φ_I, although it is the bounded summability field of a regular b-multiplicative matrix summability method.

More generally, we establish a one-one correspondence between the family of regular b-multiplicative summability methods and the family of all closed subsets of the complement N' of N in the Stone-Čech compactification βN of the (discrete) space N of positive integers. The bounded summability field of a submethod of φ_I corresponds to an open and closed subset of N'. We are unable, however, to determine which closed subsets of N' correspond to bounded summability fields of *matrix* methods of this type, although we are able to eliminate the finite subsets.

We also construct, with the aid of the continuum hypothesis, a simultaneously consistent family of submethods of φ_I which together sum every bounded sequence of real numbers.

In section 1, we define the concepts needed in this paper, and state the necessary properties of βN. In section 2, we discuss consistent families of submethods of φ_I. Section 3 is devoted to regular b-multiplicative summability methods. In section 4, we compare our simultaneously consistent family of submethods of φ_I to a pairwise b-consistent family of regular matrix summability methods which together sum every bounded real sequence that was constructed by GOFFMAN and PETERSEN [6]. Finally, in an appendix, we discuss what modifications are necessary to extend our results to complex sequences.

[1]) This paper was written while the author was an ALFRED P. SLOAN FELLOW.

I am indebted to C. GOFFMAN for acquainting me with the literature on summability, and for many valuable conversations on this subject.

1. Definitions and preliminary remarks

If X is a topological space, let $C(X)$ denote the set of all continuous real-valued functions on X, and let $C^*(X)$ denote the subset of all bounded elements of $C(X)$. Under the usual operations of pointwise addition and multiplication, these sets become algebras over the real field R. Indeed $C^*(X)$ becomes a Banach algebra if we let $\|f\| = \sup\{|f(x)| : x \in X\}$.

In particular, if N is the (discrete) space of positive integers, then $C(N)$ is the algebra of all real sequences, and $C^*(N)$ is the Banach algebra of all bounded real sequences. We use $C_0(N)$ to denote the set of all convergent real sequences.

In this paper, we capitalize on certain known properties of the Stone-Čech compactification [2], [15]. Actually, we need only consider the Stone-Čech compactification βN of N. A self-contained discussion of it is given in [14]. Proofs of many of the following assertions are given there.

(1.1). The space N is (homeomorphic to) a dense subspace of a compact (HAUSDORFF) space βN such that every $f \in C^*(N)$ has a (unique) extension $\bar{f} \in C(\beta N)$. The compact space βN, which is usually called the *Stone-Čech compactification* of N, is unique in the following sense. If Y is a compact space containing N as a dense subspace, and such that every $f \in C^*(N)$ has an extension $\bar{f} \in C(Y)$, then there is a homeomorphism of βN onto Y keeping N pointwise fixed [2], [14].

(1.2). If $Z \subset N$, let Z^β denote the closure in βN of Z. *Each Z^β is an open subset of βN.*

To see this, consider the extension $\bar{f} \in C(\beta N)$ of the bounded sequence f such that $f(n) = 0$ if $n \in Z$, and $f(n) = 1$ otherwise. Thus the complement of Z^β is closed, whence Z^β is open [3], [14].

(1.3). *If Z_1 and Z_2 are subsets of N, then $(Z_1 \cap Z_2)^\beta = Z_1^\beta \cap Z_2^\beta$. In particular, disjoint subsets of N have disjoint closures in βN* [8], [16].

(1.4). An *ultrafilter* $\mathscr{Z}$ on N is a family of nonempty subsets of N maximal with respect to the property of being closed under finite intersection. *There is a one-one correspondence between the set of ultrafilters on N and the points of βN*, obtained as follows.

If $\mathscr{Z}$ is an ultrafilter, then there is a unique point $p(\mathscr{Z}) \in \beta N$ common to every member of $\{Z^\beta : Z \in \mathscr{Z}\}$. Conversely, if $p \in \beta N$, then $\{Z \subset N : p \in Z^\beta\}$ is an ultrafilter $\mathscr{Z}$ such that $p(\mathscr{Z}) = p$.

An ultrafilter $\mathscr{Z}$ on N is called *fixed* or *free* according as $p(\mathscr{Z})$ is in N or not. Clearly, if $\mathscr{Z}$ is fixed, then it consists of all subsets of N containing $p(\mathscr{Z})$, while if $\mathscr{Z}$ is free, then every one of its elements is infinite [4], [14].

(1.5). We abbreviate the complement of N in βN by N'. Since N is open in βN, *the space N' is compact. A subset E of N' is open and closed if and only if there is a $Z \subset N$ such that $E = Z^\beta \cap N'$. Indeed, $\{Z^\beta \cap N' : Z$ is an infinite subset of $N\}$ is a base for the open subsets of N'* [14].

(1.6). Let γR denote the two-point compactification of R. Every $f \in C(N)$ is a (continuous) mapping of N into γR. Hence by a theorem of STONE, *f has a continuous extension $\hat{f}$ over βN into γR.* Clearly *if f is bounded, then* $\hat{f} = \tilde{f}$ (see [4], and [15], Theorem 88).

(1.7). If every countable intersection of neighborhoods of a point x of a compact space X is a neighborhood of x, then x is called a *P-point* of X. Equivalently, x is a P-point of X if and only if every $f \in C(X)$ is constant on a neighborhood of x. Every isolated point of X is a P-point, and every P-point with a countable base of neighborhoods is isolated. It is easily verified that not every point of an infinite compact space can be a P-point [3].

Let $\bar{g} \in C(\beta N)$ be such that $\bar{g}(n) = n^{-1}$ for all $n \in N$. Clearly $\bar{g}$ vanishes precisely on N', so no point of N' can be a P-point of βN.

On the other hand, WALTER RUDIN has shown that *if the continuum hypothesis is true, then the set of P-points of N' is a dense subset of N'.* Moreover, since N' is compact and infinite, it must also contain non-P-points [14].

(1.8). A *summability method* φ is a linear functional defined on a subspace of $C(N)$. The domain of φ is called the *summability field* $\mathcal{S}(\varphi)$ of φ, while $\mathcal{S}^*(\varphi) = C^*(N) \cap \mathcal{S}(\varphi)$ is called the *bounded summability field of* φ. If $f \in \mathcal{S}(\varphi)$, then φ is said to *sum* f.

A summability method is called *regular* if $C_0(N) \subset \mathcal{S}(\varphi)$, and $\varphi(f) = \lim_{n \to \infty} f(n)$ for all $f \in C_0(N)$.

If there is a countably infinite matrix $A = (a_{nk})$ such that for every $f \in \mathcal{S}(\varphi)$, $\varphi(f) = \lim_{n \to \infty} \sum_{k=1}^{\infty} a_{nk} f(k)$, then φ is called a *matrix summability method*, and φ is sometimes denoted by φ_A. (Henceforth, we use "matrix" to denote "countably infinite matrix".)

It is well-known that no regular matrix summability method can sum all sequences of 0's and 1's (cf., e.g., [11]).

If φ_A is a matrix summability method, Z is an infinite subset of N, and the matrix $A(Z)$ is obtained from A by deleting the n-th row of A whenever $n \notin Z$, then $\varphi_{A(Z)}$ is called a *submethod* of φ_A. Clearly $\varphi_{A(Z)}$ is regular if φ_A is regular [5].

Two summability methods φ_1, φ_2 are said to be *consistent* (resp. *b-consistent*) if whenever $f \in \mathcal{S}(\varphi_1) \cap \mathcal{S}(\varphi_2)$ [resp. $f \in \mathcal{S}^*(\varphi_1) \cap \mathcal{S}^*(\varphi_2)$], we have $\varphi_1(f) = \varphi_2(f)$. A family $\{\varphi_\alpha\}$ of summability methods is called *pairwise consistent* if every two elements of $\{\varphi_\alpha\}$ are consistent. A family $\{\varphi_\alpha\}$ of summability methods is called *simultaneously consistent* if whenever $\{\varphi_1, \ldots, \varphi_n\}$ is a finite subset of $\{\varphi_\alpha\}$, $f_i \in \mathcal{S}(\varphi_i)$, and $f_1 + \cdots + f_n = 0$, then $\varphi_1(f_1) + \cdots + \varphi_n(f_n) = 0$. Clearly every simultaneously consistent family of summability methods is pairwise consistent, but the converse does not hold. In the obvious way, we define *pairwise b-consistent* and *simultaneously b-consistent* families [10].

If φ_1, φ_2 are regular matrix summability methods such that $\mathcal{S}^*(\varphi_1) \subset \mathcal{S}^*(\varphi_2)$, it is known that φ_1 and φ_2 are b-consistent [17, p. 67.] In this case, we say that *φ_1 is weaker for bounded sequences than φ_2.*

2. Consistent families of submethods

Let I denote the identity matrix, so that φ_I denotes ordinary convergence. Clearly $f \in C_0(N)$ if and only if there is an $a \in R$ such that $\hat{f}(p) = a$ for all $p \in N'$, and $\varphi_I(f) = a$ in that event.

More generally, if $Z = \{n_1, n_2, \ldots, n_k, \ldots\}$ is an infinite subset of N, then $\lim_{k\to\infty} f(n_k)$ exists if and only if $\hat{f}$ is constant on $Z^\beta \cap N'$. Since $\varphi_{I(Z)}(f) = \lim_{n\to\infty} f(n_k)$, $\varphi_{I(Z)}(f)$ exists if and only if f is constant on $Z^\beta \cap N'$. Henceforth we abbreviate $\varphi_{I(Z)}$ by φ_Z.

Suppose that Z_1, Z_2 are infinite subsets of N. It follows easily from (1.3) that $Z_1^\beta \cap N' = Z_2^\beta \cap N'$ if and only if Z_1 and Z_2 differ only by a finite subset of N. Moreover, as was noted in (1.5), every open and closed subset of N' takes the form $Z^\beta \cap N'$ for some infinite subset Z of N. Hence we have established

(2.1). *The mapping $\varphi_Z \to Z^\beta \cap N'$ is a one-one correspondence of the family of submethods of φ_I onto the family of all open and closed subsets of N'.*

If Z_1, Z_2 are infinite subsets of N, then clearly φ_{Z_1} and φ_{Z_2} are consistent if and only if $Z_1 \cap Z_2$ is infinite. For if $Z_1 \cap Z_2$ is finite, consider any sequence f such that $f(n) = 0$ if $n \in Z_1$, and $f(n) = 1$ if $n \in Z_2$ and $n \notin Z_1 \cap Z_2$. Then $\varphi_{Z_1}(f) = 0$, while $\varphi_{Z_2}(f) = 1$. Conversely, if $Z_1 \cap Z_2$ is infinite, and both $\varphi_{Z_1}(f)$ and $\varphi_{Z_2}(f)$ exist, then $\varphi_{Z_1}(f) = \varphi_{Z_1 \cap Z_2}(f) = \varphi_{Z_2}(f)$.

Hence we have:

(2.2). If $\{\varphi_{Z_\alpha}\}$ is a simultaneously *consistent family of submethods of φ_I, then it is a subfamily of $\{\varphi_Z : Z \in \mathscr{Z}\}$, where $\mathscr{Z}$ is a free ultrafilter on N* (1.4).

(2.3). If $\mathscr{Z}$ is a free ultrafilter on N, then as was noted in (1.4), there is a unique point $p \in N'$ common to all Z^β such that $Z \in \mathscr{Z}$. Moreover, it is clear from (2.1) and the preceeding comments that *if $Z \in \mathscr{Z}$, and $f \in \mathscr{S}(\varphi_Z)$, then* $\varphi_Z(f) = \hat{f}(p)$.

Suppose that $\mathscr{Z}$ is a free ultrafilter on N. If Z is a subset of N that meets every element of $\mathscr{Z}$, then by definition of ultrafilter (1.4), $Z \in \mathscr{Z}$. So, we have

(2.4). *If $\mathscr{Z}$ is a free ultrafilter on N then $\{\varphi_Z : Z \in \mathscr{Z}\}$ is a simultaneously consistent family of submethods of φ_I maximal with respect to the property of being pairwise consistent.*

We have shown above that for any free ultrafilter $\mathscr{Z}$, and any $Z \in \mathscr{Z}$, that $f \in \mathscr{S}(\varphi_Z)$ if and only if $\hat{f}$ is constant on $Z^\beta \cap N'$. Hence $f \in \mathscr{S}(\varphi_Z)$ for some $Z \in \mathscr{Z}$ if and only if $\hat{f}$ is constant on a neighborhood in N' of $p(\mathscr{Z})$ (1.5). Hence we have,

(2.5). *If $\mathscr{Z}$ is a free ultrafilter on N, then every $f \in C^*(N)$ is summed by some φ_Z with $Z \in \mathscr{Z}$ if and only if $p(\mathscr{Z})$ is a P-point of N'* (1.7).

Next we state the main result of this section.

(2.6) THEOREM. *Suppose that the continuum hypothesis is true. If φ_A is any regular matrix summability method, then there is a simultaneously consistent family of submethods of φ_A such that every $f \in C^*(N)$ is summed by at least one member of the family.*

Proof. It suffices to prove the theorem in case $A = I$, for if $\{\varphi_{Z_\alpha}\}$ is such a family of submethods of φ_I, then $\{\varphi_{I(Z_\alpha),A}\}$ is the desired family of submethods of φ_A.

The validity of the continuum hypothesis implies the existence of a P-point p of N' by (1.7). By (1.4), there is a free ultrafilter $\mathscr{Z}$ on N with $p(\mathscr{Z}) = p$. By (2.4) and (2.5) $\{\varphi_Z : Z \in \mathscr{Z}\}$ is the desired family.

(2.7). REMARK. The technique employed in the first paragraph of the proof shows that (2.2), (2.4), and (2.5) are valid if φ_I is replaced by any regular matrix summability method.

(2.8). REMARK. If φ_A is a regular row-finite matrix summability method, we will say that $f \in C(N)$ is summed by φ_A *in the extended sense* in case $\varphi_A(f)$ exists, or if $\lim_{n\to\infty} \sum_{k=1}^{\infty} a_{nk} f(k) = \pm\infty$, where $A = (a_{nk})$. It follows easily from (1.6) and the above that if $\mathscr{Z}$ is a free ultrafilter such that $p(\mathscr{Z})$ is a P-point, then for every $f \in C(N)$, there is a $Z \in \mathscr{Z}$ such that f is summed in the extended sense by $\varphi_{A(Z)}$.

3. Multiplicative summability methods

A summability method φ will be called *multiplicative* (resp. *b-multiplicative*) if whenever $f, g \in \mathscr{S}(\varphi)$ [resp. $f, g \in \mathscr{S}^*(\varphi)$], we have $\varphi(fg) = \varphi(f)\,\varphi(g)$, and if $\mathscr{S}^*(\varphi)$ is uniformly closed.

By a well known theorem of MAZUR and ORLICZ [*11*] (see also [*17*, p. 71]), a regular matrix summability method is multiplicative if and only if it is a submethod of φ_1.

In [*1*], it was conjectured that every regular b-multiplicative matrix summability method is equivalent *for bounded sequences* to a submethod of φ_1. Below we will show that this conjecture is false. First, however, we will study regular b-multiplicative summability methods in general.

Let φ denote any regular b-multiplicative summability method. Then $\mathscr{S}^*(\varphi)$ is a uniformly closed subalgebra of the Banach algebra $C^*(N)$. It follows easily from (1.1), that $C^*(N)$ is isomorphic as a Banach algebra to $C(\beta N)$.

If $p, q \in \beta N$, let pRq denote the assertion that $\tilde{f}(p) = \tilde{f}(q)$ for all $f \in \mathscr{S}^*(\varphi)$. It is clear that R is an equivalence relation on βN, and that distinct points of N lie in distinct equivalence classes. Let X denote the quotient space of X by the relation R. It is clear that X is a compact (HAUSDORFF) space (cf., e.g., [*9*, Chapter 3]). Let δ denote the quotient mapping of βN onto X. If $f \in \mathscr{S}^*(\varphi)$, define the function f' on X by letting $f'(\delta(p)) = \tilde{f}(p)$ for all $p \in \beta N$. It is easily seen that the mapping $f \to f'$ is an isomorphism of $\mathscr{S}^*(\varphi)$ into the Banach algebra $C(X)$ (with the usual "sup" norm). Moreover the image of $\mathscr{S}^*(\varphi)$ under this isomorphism is a uniformly closed subalgebra of $C(X)$ that contains the constant function, and such that if x_1 and x_2 are distinct points of X, there is an $f \in \mathscr{S}^*(\varphi)$ such that $f'(x_1) \neq f'(x_2)$. Hence, by the celebrated Stone-Weierstrass approximation theorem [*9*, p. 244], the isomorphism is onto.

Clearly φ induces a multiplicative linear function φ' on the Banach algebra $C(X)$, so there is a point $x_0 \in X$ such that for any $f' \in C(X)$, $\varphi'(f') = f'(x_0)$.

Hence $f \in \mathscr{S}^*(\varphi)$ if and only if f is constant on the closed set $\delta^{-1}(x_0)$. We denote this closed set by $F(\varphi)$. Clearly $F(\varphi) \subset N'$.

Conversely, if F is any closed subset of N', then $\{f \in C^*(N) : \bar{f}$ is constant on $F\}$ is the bounded summability field of a regular b-multiplicative summability method φ such that $F = F(\varphi)$.

We summarize the above with the following theorem.

(3.1). THEOREM. *The mapping $\varphi | C^*(N)$ is a one-one correspondence between the family of restrictions to $C^*(N)$ of regular b-multiplicative summability methods and the family of all closed subsets of N'. Moreover $f \in \mathscr{S}^*(\varphi)$ if and only if f is constant on $F(\varphi)$.*

If φ_1, φ_2 are two regular b-multiplicative summability methods such that $F(\varphi_1) \subset F(\varphi_2)$, then $\mathscr{S}^*(\varphi_2) \subset \mathscr{S}^*(\varphi_1)$. Conversely, if $\mathscr{S}^*(\varphi_2) \subset \mathscr{S}^*(\varphi_1)$, and $F(\varphi_1)$ has more then one point, then $F(\varphi_1) \subset F(\varphi_2)$. For, otherwise, there is a p in $F(\varphi_1)$, but not in $F(\varphi_2)$ and a $q \neq p$ in $F(\varphi_1)$. Choose an $f \in C^*(N)$ such that $\bar{f}(p) = 0$ and $\bar{f}(x) = 1$ if $x \in F(\varphi_2) \cup \{q\}$. Then $\bar{f}$ is constant on $F(\varphi_2)$, but is not constant on $F(\varphi_1)$, so $f \in \mathscr{S}^*(\varphi_2)$, and $f \notin \mathscr{S}^*(\varphi_1)$. Hence we have shown

(3.2). *If φ_1, φ_2 are regular b-multiplicative summability methods such that $F(\varphi_1)$ contains more than one point, then $F(\varphi_1) \subset F(\varphi_2)$ if and only if $\mathscr{S}^*(\varphi_2) \subset \mathscr{S}^*(\varphi_1)$. Hence if φ is a regular b-multiplicative summability method that fails to sum a bounded divergent sequence, then the restriction of φ to $C^*(N)$ is uniquely determined by $\mathscr{S}^*(\varphi)$.*

As was noted in (1.8), the conclusion of the last part of (3.2) holds for any regular *matrix* summability method (whether b-multiplicative or not).

It seems natural to ask which closed subsets F of N' are such that there is a regular b-multiplicative *matrix* summability method such that $F = F(\varphi)$? Theorem 2.1 tells us that every open and closed subset F of N' has this property, and that every such F determines a submethod of φ_I. We show next the existence of regular b-multiplicative matrix methods φ such that $F(\varphi)$ is not open and closed.

(3.3). EXAMPLE. Let $Z_1, Z_2, \ldots, Z_n, \ldots$ denote a sequence of subsets of N such that for each $n \in N$, $Z_n \subset Z_{n+1}$, and there are infinitely many elements of Z_{n+1} that are not in Z_n, and let $F_n = Z_n^\beta \cap N'$. It is clear that $F(\varphi_{Z_n}) = F_n$ for each $n \in N$, and that $F_n \subset F_{n+1}$. So, by (3.2), $\mathscr{S}^*(\varphi_{Z_n}) \supset \mathscr{S}^*(\varphi_{Z_{n+1}})$ for each $n \in N$.

A well known theorem of Broudno states that the intersection of a countable decreasing chain of bounded summability fields of regular matrix summability methods is the bounded summability field of a regular matrix summability method [17, p. 54]. Hence there is a regular matrix A such that $\mathscr{S}^*(\varphi_A) = \bigcap_{n=1}^{\infty} \mathscr{S}^*(\varphi_{Z_n})$. Moreover it is clear that φ_A is b-multiplicative, and it follows easily from (3.2) that $F(\varphi_A)$ is the closure in N' of $\bigcup_{n=1}^{\infty} F_n$.

As we have already seen, φ_A is equivalent for bounded sequences to a submethod of φ_I if and only if $F(\varphi_A)$ is open and closed. If $F(\varphi_A) = F$ were open and closed, then by (1.5), there would be an infinite subset Z of N such that $F = Z^\beta \cap N'$. Since $F_n \subset F$ for each $n \in N$, Z contains all but a finite number

of the elements of Z_n. Hence for each $n \in N$, there are infinitely many elements of $Z_{n+1} \cap Z$ that are not in $Z_n \cap Z$; choose any such and call it x_n. Let T denote the set obtained from Z by deleting $\{x_n : n \in N\}$. Now for each $n \in N$, $T \cap Z_n$ differs from Z_n by only a finite number of elements, so $(T \cap Z_n)^\beta \cap N' = Z_n^\beta \cap N' = F_n$. It follows that $F \subset T^\beta \cap N'$. On the other hand, there are infinitely many elements of Z that are not in T, so $T^\beta \cap N'$ is a *proper* subset of $Z^\beta \cap N' = F$. This contradiction shows that $F(\varphi_A)$ is not open and closed, and hence that *the regular b-multiplicative matrix summability method φ_A is not equivalent for bounded sequences to a submethod of φ_I.*

Actually, what we have given above is a proof of the known fact that the Boolean algebra of open and closed subsets of N' is not σ-complete (see [7]).

If the reader has the patience to do so, he can construct an explicit matrix A satisfying the conditions of Example 3.3 by interweaving with proper weighting factors the rows of the matrices $I(Z_n)$, $n = 1, 2, \ldots$, as in BROUDNO's proof of the existence of a matrix A such that $\mathscr{S}^*(\varphi_A) = \bigcap_{n=1}^{\infty} \mathscr{S}^*(\varphi_{Z_n})$ (see [17, p. 54]).

Note also that the argument given above cannot be used to conclude (falsely) that φ_A is a multiplicative-summability method, even though each φ_{Z_n} is. For, a countable intersection of a strictly decreasing chain of summability fields of regular matrix summability methods need not be the summability field of a regular matrix summability method [17, p. 52].

The method φ_A is, however, weaker for bounded sequences than a submethod of φ_I (1.8).

By way of positive results in this direction, we have only

(3.4). THEOREM. *If φ_A is a regular b-multiplicative matrix summability method, then $F(\varphi_A)$ has power 2^c (where c is the cardinal number of the continuum).*

PROOF. It suffices to prove that $F(\varphi_A)$ is infinite, for every infinite closed subset of βN has power 2^c [12].

Suppose that there is a positive integer k such that $F(\varphi_A)$ has precisely k elements. If $k = 1$, then $\mathscr{S}^*(\varphi_A) = C^*(N)$, which is impossible (1.8), so we may assume that $k > 1$. Suppose that $F(\varphi_A) = \{p_1, \ldots, p_k\}$, so that φ_A sums $\{f \in C^*(N) : \bar{f}(p_1) = \cdots = \bar{f}(p_k)\}$. There is a $g \in C^*(N)$ such that $\bar{g}(p_1) = 0$, while $\bar{g}(p_2) = \cdots = \bar{g}(p_k) = 1$, and there is a submethod $A_1 = A(Z)$ that sums g. Moreover, since $A(Z) = I(Z)A$, $\mathscr{S}^*(\varphi_A) \subset \mathscr{S}^*(\varphi_{A_1})$. Suppose that $h \in C^*(N)$, and $\bar{h}(p_1) \neq \bar{h}(p_2) = \cdots = \bar{h}(p_k)$. Let $h_1 = (h - \bar{h}(p_1))(\bar{h}(p_2) - \bar{h}(p_1))^{-1}$. Then $(\bar{h}_1 - \bar{g})(p_1) = 0$, and $(\bar{h}_1 - \bar{g})(p_2) = \cdots = (\bar{h} - \bar{g})(p_k) = 1$, so $(h_1 - g) \in \mathscr{S}^*(\varphi_{A_1})$. Thus, $h_1 = g + (h_1 - g)$ is in $\mathscr{S}^*(\varphi_{A_1})$, whence $h = (\bar{h}(p_2) - \bar{h}(p_1))\, h_1 + \bar{h}(p_1)$ is in $\mathscr{S}^*(\varphi_{A_1})$. It follows that $\mathscr{S}^*(\varphi_{A_1})$ contains $\{f \in C^*(N) : \bar{f}(p_2) = \cdots = \bar{f}(p_k)\}$. Continuing in this way, we can construct a submethod φ_{A_k} of φ_A that sums all bounded sequences. This contradiction yields the theorem.

4. Some remarks on a paper of GOFFMAN and PETERSEN

Let $\mathfrak{A}$ denote the class of all positive regular matrices such that for every $A \in \mathfrak{A}$, the number $\frac{1}{2}$ appears at least once in each row and each column of A. In [6], GOFFMAN and PETERSEN show that $\{\varphi_A : A \in \mathfrak{A}\}$ is a pairwise b-con-

sistent family of regular matrix summability methods such that every bounded sequence f is summed by at least one member φ_A of the family. Indeed, if $U(f)$, $l(f)$ denote respectively the limit superior and limit inferior of the sequence f, then $\varphi_A(f) = \frac{U(f)+l(f)}{2}$. Thus, the family A defines a functional Φ on $C^*(N)$ whose value at f is $\frac{U(f)+l(f)}{2}$.

Unfortunately, (as was pointed out to the author by C. GOFFMAN), Φ is not a summability method, i.e., Φ is not linear. Equivalently, $\{\varphi_A : A \in \mathfrak{A}\}$ is not simultaneously b-consistent. For, let f_1, f_2, f_3 denote respectively the sequences $\{1, 0, 0, 1, 0, 0, \ldots\}$, $\{0, 1, 0, 0, 1, 0, \ldots\}$, $\{-1, -1, 0, -1, -1, 0, \ldots\}$. Then $\Phi(f_1+f_2+f_3) = \Phi(0) = 0$, while $\Phi(f_1)+\Phi(f_2)+\Phi(f_3) = \frac{1}{2}+\frac{1}{2}-\frac{1}{2} = \frac{1}{2}$. Moreover, as was pointed out by AGNEW in hiw review of [6] (Mathematical Reviews, vol. 17, p. 1200), the family $\{\varphi_A : A \in \mathfrak{A}\}$ is not pairwise consistent (although, of course, it is pairwise b-consistent).

On the other hand, if $\mathscr{Z}$ is an ultrafilter on N such that $p(\mathscr{Z})$ is a P-point of N'. Then, as is shown in section 2, $\{\varphi_Z : Z \in \mathscr{Z}\}$ is a simultaneously consistent family of regular matrix summability methods such that every bounded sequence is summed by some φ_Z. Moreover, the functional $\Phi_{\mathscr{Z}}$ defined on $C^*(N)$ by this family [i.e., if $f \in C^*(N)$, let $\Phi_{\mathscr{Z}}(f) = f(p(\mathscr{Z}))$] is both linear and continuous.

Hence this latter family has a number of properties not possessed by the one constructed by GOFFMAN and PETERSEN. On the other hand, the proof of its existence relies on the continuum hypothesis (2.6), and it is more difficult to visualize than that of GOFFMAN and PETERSEN.

In [6], these authors pose the following problem. Can every pairwise b-consistent family of regular matrix summability methods be enlarged to a pairwise b-consistent family of regular matrix summability methods such that every bounded sequence is summed by some member of the enlarged family? In [13], PETERSEN answers this question in the affirmative for countable family of methods with bounded norm. (See [13] for the definition of norm.)

If $\mathscr{Z}$ is an ultrafilter on N such that $p(\mathscr{Z})$ is not a P-point, then by (2.5) the simultaneously consistent family $\{\varphi_A : Z \in \mathscr{Z}\}$ does not sum every bounded sequence. (It is easy to show, however, that it will sum all bounded sequences with only a finite number of limit points.) This family impresses the author as a likely candidate for a negative answer to this problem, but he has been unable to discover a proof in either direction.

Appendix

All of the above discussion has been concerned with sequences of real numbers. It seems natural to ask what can be done with sequences of complex numbers.

All of the results of section 2 can be extended with trivial modification to handle complex sequences as well. For, every bounded complex-valued function on N has a (unique) continuous extension over βN.

The results of section 3, however, rely on the Stone-Weierstrass approximation theorem. These too, can be extended to handle complex sequences if we require that whenever a function on N is summed by one of our summability methods φ, then φ also sums its complex conjugate.

Added October 9, 1958. The author had conjectured that every regular b-multiplicative matrix summability method is weaker for bounded sequences than a submethod of φ_1. At the International Congress of Mathematicians in Edinburgh, P. ERDÖS, relaying a message from K. ZELLER, pointed out that this is false. See reference ZELLER, 53e) in [17]. The author will comment on this example and its ramifications in a future note.

References

[1] BRAUER, G.: Evaluation of product sequences by matrix methods. Amer. Math. Monthly **63**, 323–326 (1956). — [2] ČECH, E.: On bicompact spaces. Ann. of Math. **38**, 823–844 (1937). — [3] GILLMAN, L., and M. HENRIKSEN: Concerning rings of continuous functions. Trans. Amer. Math. Soc. **77**, 340–362 (1954). — [4] GILLMAN, L., M. HENRIKSEN and M. JERISON: On a theorem of Gelfand and Kolmogoroff concerning maximal ideals in rings of continuous functions. Proc. Amer. Math. Soc. **5**, 445–455 (1954). — [5] GOFFMAN, C., and G. PETERSEN: Submethods of regular matrix summability methods. Canad. J. Math. **8**, 40–46 (1956). — [6] GOFFMAN, C., and G. PETERSEN: Consistent limitation methods. Proc. Amer. Math. Soc. **7**, 367–369 (1956). — [7] HENRIKSEN, M.: Some remarks on a paper of ARONSZAJN and PANITCHPAKDI. Pacific J. Math. **7**, 1619 to 1621 (1957). — [8] KATETOV, M.: A theorem on the Lebesgue dimension. Časopis Mat. Fys. **75**, 79–87 (1950). — [9] KELLEY, J.: General topology. New York: Van Nostrand 1955. — [10] LORENTZ, G., and K. ZELLER: Über Paare von Limitierungsverfahren. Math. Z. **68**, 422–438 (1957). — [11] MAZUR, S., and W. ORLICZ: On linear methods of summability. Studia Math. **14**, 129–160 (1954). — [12] NOVÁK, J.: On the cartesian product of two compact spaces. Fund. Math. **40**, 106–112 (1953). — [13] PETERSEN, G.: Sets of consistent summation methods. J. London Math. Soc. **32**, 377–379 (1957). — [14] RUDIN, W.: Homogeneity problems in the theory of Čech compactifications. Duke Math. J. **23**, 409–420 (1956). — [15] STONE, M. H.: Applications of the theory of Boolean rings to general topology. Trans. Amer. Math. Soc. **41**, 375–481 (1937). — [16] WALLACE, A.: Extensional invariance. Trans. Amer. Math. Soc. **70**, 97–102 (1951). — [17] ZELLER, K.: Theorie der Limitierungsverfahren. Ergebnisse der Mathematik und ihrer Grenzgebiete. Berlin: Springer 1958.

Dept. of Math., Purdue University, Lafayette, Indiana (U.S.A.)

(Eingegangen am 11. August 1958)

WALTER, W.
Math. Zeitschr. 71, 436–453 (1959)

Über die Differentialgleichung $u_{xy}=f(x,y,u,u_x,u_y)$

II. Existenzsätze für das charakteristische Anfangswertproblem[1])

Von
WOLFGANG WALTER

§ 6. Vorbemerkungen

Wir behandeln auch in diesem zweiten Teil nur das charakteristische Anfangswertproblem für die Differentialgleichung

$$u_{xy}=f(x,y,u,u_x,u_y), \tag{1}$$

wie es in § 1 genauer beschrieben wurde; wir übernehmen die Bezeichnungen aus § 1. Die Funktion $f(x,y,u,v,w)$ sei im folgenden, solange nicht ausdrücklich etwas anderes gesagt wird, *stetig und beschränkt* für $(x,y)\in R$ und $-\infty<u,v,w<\infty$.

Man weiß seit PICARD [1], daß eine Lösung dieser Aufgabe existiert, sobald $f(x,y,u,v,w)$ in den Variablen u,v,w einer Lipschitz-Bedingung genügt. In einer Reihe von Arbeiten wurde dieses Ergebnis verschärft. Wir nennen hier unter den neueren Arbeiten nur HARTMAN-WINTNER [8], ALEXIEWICZ-ORLICZ [9] und LEEHEY [11] und verweisen auf die Arbeit von DIAZ [10], wo eine übersichtliche Zusammenstellung der bekannten Existenz- und Eindeutigkeitssätze und ausführliche Literaturhinweise gegeben werden. Der allgemeinste bekannte Existenzsatz von LEEHEY [11] und HARTMAN-WINTNER [8] sagt aus, daß man eine Lipschitz-Bedingung nur in v und w zu fordern braucht, d. h. also, daß eine Abschätzung

$$|f(x,y,u,v,w)-f(x,y,u,\bar v,\bar w)|\leqq L(|v-\bar v|+|w-\bar w|)$$

hinreichend für die Existenz einer Lösung ist. Von großer Wichtigkeit für die Theorie ist ferner das von HARTMAN-WINTNER [8] gegebene Beispiel einer stetigen und beschränkten Funktion f, zu der es keine Lösung $u(x,y)\in C^*$ der Differentialgleichung (1) gibt. Damit war gezeigt, daß die in anderen Teilen der Theorie, besonders bei den Eindeutigkeitssätzen sichtbare weitgehende Analogie zum Anfangswertproblem für die *gewöhnliche* Differentialgleichung

$$y'(x)=f(x,y),\qquad y(0)=\eta \tag{2}$$

[1]) Diese Arbeit wurde teilweise unterstützt von der United States Air Force [Air Force Office of Scientific Research-Contract AF 49 (638)–228]. Sie besteht aus drei Teilen. Teil I ist im vorangehenden Heft dieser Zeitschrift erschienen, Teil III erscheint im folgenden Heft. Formeln und Sätze sind in jedem Teil für sich durchnumeriert. Bei Verweisen bedeutet z. B. (I.17) die Formel (17) von Teil I, Satz I.5 bedeutet Satz 5 von Teil I.

beim Existenzproblem zu bestehen aufhört: der Existenzsatz von PEANO hat kein Analogon. Es erhebt sich dann die Frage, wo die Grenze zwischen Lösbarkeit und Nicht-Lösbarkeit liegt, und im besonderen, ob unter den allgemeinen Eindeutigkeitsbedingungen von Teil I eine Lösung überhaupt existiert.

Im nächsten Paragraphen werden wir einen allgemeinen Existenzsatz beweisen. Danach ist es für die Existenz hinreichend, wenn $f(x, y, u, v, w)$ in der Variablen w einer Abschätzung

$$|f(x, y, u, v, w) - f(x, y, u, v, \overline{w})| \leqq g(x, |w - \overline{w}|) \tag{3}$$

sowie einer entsprechenden Abschätzung in v genügt. Die Funktion $g(x, z)$ ist von der Art, wie sie in dem allgemeinen Eindeutigkeitssatz von KAMKE [7, S. 139] für die *gewöhnliche* Differentialgleichung (2) vorkommt (s. auch § 2). Die genauen Voraussetzungen werden im nächsten Paragraphen gegeben; zulässig sind z. B. die Funktionen $g = K(x)\, z$ (Lipschitz-Bedingung), $g = K(x)\omega(z)$ (Bedingung von OSGOOD und MONTEL; Voraussetzungen wie bei Beispiel (a), (b) in § 2) und $g = z\left[\frac{\beta}{x} + K(x)\right]$, $0 < \beta < 1$[2]). Die Bedingung (3) ist eine Eindeutigkeitsbedingung für die *gewöhnliche* Differentialgleichung

$$w'(x) = f(x, y, u, v, w(x)) \qquad (y, u, v \text{ feste Parameter}). \tag{4}$$

Der § 8 enthält einige Verallgemeinerungen des Existenzsatzes sowie Beispiele von Funktionen f, für welche das Anfangswertproblem im allgemeinen nicht lösbar ist. Das oben zitierte, übrigens nicht ganz einfache Beispiel von HARTMAN-WINTNER hat die Eigenschaft, daß die zugehörige Gl. (1) überhaupt keine Lösung zuläßt. Wir geben dagegen einfache Beispiele von Funktionen f und zugehörigen Anfangswerten σ und τ, für die die Anfangswertaufgabe keine Lösung hat. Diese Beispiele zeigen in Verbindung mit dem Existenzsatz den engen Zusammenhang zwischen dem *Existenzproblem für die charakteristische Anfangswertaufgabe* der Gl. (1) und dem *Eindeutigkeitsproblem für die gewöhnliche Differentialgleichung* (4).

Die Untersuchung der Frage, wann das Verfahren der sukzessiven Approximation konvergiert, bildet den Inhalt von § 9. Die Beweise haben große Ähnlichkeit mit den Eindeutigkeitsbeweisen von Teil I und sind deshalb nicht immer ausführlich dargestellt. Die Ergebnisse können als Verallgemeinerung der Approximationssätze, welche NAGUMO [6] für die gewöhnliche Differentialgleichung (2) aufgestellt hat, angesehen werden.

Der Existenzbeweis des folgenden Paragraphen verläuft in der üblichen Weise. Man konstruiert zunächst eine Folge von „Näherungslösungen" und zeigt, daß diese gleichgradig stetig sind. Nach dem Satz von ASCOLI existiert dann eine gleichmäßig konvergente Teilfolge, und es bleibt nur noch nachzuweisen, daß deren Limes eine Lösung der Aufgabe darstellt. Der schwierigste Schritt ist hierbei immer der Nachweis der gleichgradigen Stetigkeit; er hängt

[2]) Für $K \equiv 0$ ist das die Bedingung von ROSENBLATT [14], für $\beta = 0$ die Lipschitz-Bedingung. Man beachte jedoch, daß $\beta = 1$ (Bedingung von NAGUMO) nicht zugelassen ist.

entscheidend davon ab, in welcher Weise die Näherungslösungen konstruiert sind.

Unser Konstruktionsverfahren soll zuerst am Beispiel der gewöhnlichen Differentialgleichung

$$y' = f(x, y) \quad (0 \leqq x \leqq a), \qquad y(0) = \eta \tag{2}$$

verdeutlicht werden. In diesem eindimensionalen Fall ist es mit dem von TONELLI [12] angegebenen Verfahren verwandt (dieses Verfahren wurde von CONTI [15] zum Beweis eines Existenzsatzes für das charakteristische Anfangswertproblem der Gl. (1) benutzt, bei dem vorausgesetzt wird, daß f einer Lipschitz-Bedingung in v und w und die Anfangswerte $\sigma(x)$, $\tau(y)$ einer Lipschitz-Bedingung in x bzw. y genügen).

Zu jedem $\alpha > 0$ gibt es eine „Näherungslösung" $y_\alpha(x)$ mit den Eigenschaften

$$y_\alpha(x) = \begin{cases} \eta & \text{für } x \leqq 0 \\ \eta + \int_0^x f(t, y_\alpha(t-\alpha))\, dt & \text{für } 0 < x \leqq a. \end{cases} \tag{5}$$

Denn ist zunächst $0 < x \leqq \alpha$, so steht in der zweiten Zeile von (5) ein wohldefiniertes Integral, da $t - \alpha \leqq x - \alpha \leqq 0$ und deshalb $y_\alpha(t-\alpha) = \eta$ ist; die Funktion y_α ist damit für $x \leqq \alpha$ definiert. Nun sei $\alpha < x \leqq 2\alpha$. Man sieht sofort, daß auch jetzt im Integral in (5) nur bekannte Größen stehen, also y_α für $x \leqq 2\alpha$ eindeutig definiert werden kann. Auf diese Weise fortfahrend, erhält man nach endlich vielen Schritten eine Funktion y_α, welche für alle $x \leqq a$ definiert ist und die Gl. (5) befriedigt. Mit Hilfe dieser Näherungslösungen läßt sich der Existenzsatz von PEANO für die Gl. (2) sehr leicht beweisen. Ist die Funktion $f(x, y)$ beschränkt, so sind die Funktionen $y_\alpha(x)$ gleichgradig stetig. Nach dem Satz von ASCOLI gibt es dann eine Nullfolge α_n derart, daß die Folge der zugehörigen Näherungslösungen y_{α_n} gleichmäßig gegen eine stetige Funktion $y(x)$ konvergiert. Der Grenzübergang $n \to \infty$ kann in (5) unter dem Integralzeichen vorgenommen werden, d.h. die Funktion $y(x)$ ist eine Lösung von (5) mit $\alpha = 0$ und damit auch Lösung des Anfangswertproblems (2).

Dieser Beweis ist ohne Änderung auch für den allgemeineren Fall, daß $f(x, y)$ in x meßbar, in y stetig und $|f(x, y)| \leqq M(x)$ mit einer über $[0, a]$ summierbaren Funktion $M(x)$ ist (Existenzsatz von CARATHEODORY), durchführbar. Desgleichen kann man, worauf wir hier nicht näher eingehen wollen, auf diese Weise auch Lösungen von nichtlinearen Integralgleichungen vom Volterra-Typus erhalten. Die Übertragung auf den Fall, daß $y(x)$ eine Vektorfunktion ist, liegt ebenfalls auf der Hand. Im nächsten Paragraphen soll diese Methode in geeigneter Modifikation auf das vorliegende Problem angewandt werden.

§ 7. Der Existenzsatz

Wir schicken den einfachen Fall der Differentialgleichung

$$u_{xy} = f(x, y, u) \tag{6}$$

voraus. Gleichwertig mit der charakteristischen Anfangswertaufgabe ist die Aufgabe, eine Lösung der Integralgleichung

$$u(x,y)=\sigma(x)+\tau(y)-\sigma(0)+\int_0^x\int_0^y f(s,t,u)\,ds\,dt \tag{7}$$

zu finden (s. (I.2)[1])). Es ist bekannt, daß sie für jede stetige, beschränkte Funktion $f(x, y, u)$ lösbar ist. Die Lösung ist im allgemeinen nicht eindeutig. Mit der im vorigen Paragraphen beschriebenen Methode erhält man eine Lösung auf folgendem Wege. Zu jedem $\alpha>0$ bestimmt man zunächst eine Näherungslösung $u_\alpha(x, y)$, die der Gleichung

$$u_\alpha(x,y)=\begin{cases}\tau(y) & \text{für } x\leqq 0,\ 0\leqq y\leqq b\\ \tau(y)+\sigma(x)-\sigma(0)+\int_0^x\int_0^y f(s,t,u_\alpha(s-\alpha,t))\,ds\,dt & \text{für } (x,y)\in R\end{cases} \tag{8}$$

genügt. Jede Funktion $u_\alpha(x, y)$ ist also für $x\leqq a$, $0\leqq y\leqq b$ definiert, und man erhält sie ähnlich wie im vorigen Paragraphen zunächst für $0\leqq x\leqq\alpha$, dann für $\alpha<x\leqq 2\alpha, \ldots$. Aus der Beschränktheit von $f(x, y, u)$ folgt wieder die gleichgradige Stetigkeit aller u_α. Denn ist $(x, y)\in R$ und $(x+h, y+k)\in R$ und etwa $|f|\leqq M$, so ist

$$\begin{aligned}|u_\alpha(x+h,y+k)-u_\alpha(x,y)|\leqq{}&|\sigma(x+h)-\sigma(x)|+\\&+|\tau(y+k)-\tau(y)|+M(a|k|+b|h|)\end{aligned} \tag{9}$$

für alle $\alpha>0$. Es existiert also eine gleichmäßig konvergente Folge $u_{\alpha_n}(x, y)$ mit $\alpha_n\to 0$. Ist $u(x, y)$ ihr Grenzwert, so folgt aus (8) für $\alpha=\alpha_n\to 0$ die Gl. (7), d. h. $u(x, y)$ ist eine Lösung der charakteristischen Anfangswertaufgabe.

Die Auszeichnung der x-Koordinate bei der Näherungsfunktion u_α ist willkürlich. Man hätte ebensogut das Integral in (8) durch eines der Integrale

$$\int_0^x\int_0^y f(s,t,u_\alpha(s,t-\alpha))\,ds\,dt \quad \text{oder} \quad \int_0^x\int_0^y f(s,t,u_\alpha(s-\alpha,t-\alpha))\,ds\,dt \tag{10}$$

ersetzen können. Es wäre dann allerdings nötig, u_α für $y<0$ irgendwie in stetiger Weise zu definieren.

Wenden wir uns nun der allgemeinen Differentialgleichung (1) zu. Es ist leicht zu sehen, daß eine äquivalente Formulierung des charakteristischen Anfangsproblems durch das folgende System von Integralgleichungen für drei Funktionen u, v, w

$$\begin{cases}u(x,y)=\sigma(x)+\tau(y)-\sigma(0)+\int_0^x\int_0^y f(s,t,u,v,w)\,ds\,dt\\ v(x,y)=\sigma'(x)+\int_0^y f(x,t,u,v,w)\,dt\\ w(x,y)=\tau'(y)+\int_0^x f(s,y,u,v,w)\,ds\end{cases} \tag{11}$$

gegeben ist. Zur Lösung von (11) konstruieren wir Näherungsfunktionen $u_\alpha, v_\alpha, w_\alpha$, die dem System

$$(12)\quad \begin{cases} u_\alpha(x,y) = \sigma(x) + \tau(y) - \sigma(0) + \\ \qquad + \int_0^x \int_0^y f\big(s,t,u_\alpha(s-\alpha,t), v_\alpha(s,t-\alpha), w_\alpha(s-\alpha,t)\big)\, ds\, dt \\ v_\alpha(x,y) = \sigma'(x) + \int_0^y f\big(x,t,u_\alpha(x-\alpha,t), v_\alpha(x,t-\alpha), w_\alpha(x-\alpha,t)\big)\, dt \\ w_\alpha(x,y) = \tau'(y) + \int_0^x f\big(s,y,u_\alpha(s-\alpha,y), v_\alpha(s,y-\alpha), w_\alpha(s-\alpha,y)\big)\, ds \end{cases}$$

für $(x,y) \in R$ genügen sollen[2a]). Wir setzen für alle $\alpha > 0$

$$(13)\quad \begin{cases} u_\alpha(x,y) = \tau(y) & \text{für } x \leqq 0,\ 0 \leqq y \leqq b \\ v_\alpha(x,y) = \sigma'(x) & \text{für } 0 \leqq x \leqq a,\ y \leqq 0 \\ w_\alpha(x,y) = \tau'(y) & \text{für } x \leqq 0,\ 0 \leqq y \leqq b. \end{cases}$$

Zunächst sei $0 \leqq x \leqq \alpha$, $0 \leqq y \leqq \alpha$. Dann läßt sich $u_\alpha, v_\alpha, w_\alpha$ durch (12) bestimmen, da in den Integralen nur Funktionswerte vorkommen, die gemäß (13) schon definiert sind. Mit Hilfe der so gewonnenen Werte lassen sich die Funktionen $u_\alpha, v_\alpha, w_\alpha$ jetzt auch für das Quadrat $\alpha \leqq x \leqq 2\alpha$, $0 \leqq y \leqq \alpha$ bestimmen. Nach endlich vielen Schritten erhält man so einen „Streifen" $0 \leqq x \leqq a$, $0 \leqq y \leqq \alpha$, in welchem die Näherungsfunktionen definiert sind und von dem aus man sie in den zweiten Streifen $0 \leqq x \leqq a$, $\alpha \leqq y \leqq 2\alpha$ fortsetzen kann, usw. Kurz gesagt, man überzieht das Rechteck R mit Quadraten von der Seitenlänge α und bestimmt die Näherungsfunktionen $u_\alpha, v_\alpha, w_\alpha$, indem man, in der linken unteren Ecke beginnend, von Quadrat zu Quadrat fortschreitet (dabei kann man auch eine andere Reihenfolge als die eben beschriebene wählen). Nach (12) und (13) sind die so erhaltenen Funktionen $u_\alpha, v_\alpha, w_\alpha$ in ihren Definitionsbereichen $x \leqq a$, $0 \leqq y \leqq b$ bzw. $0 \leqq x \leqq a$, $y \leqq b$ bzw. $x \leqq a$, $0 \leqq y \leqq b$ stetig.

Wenn die auf diese Weise konstruierten Näherungsfunktionen gleichgradig stetig sind, so erlaubt uns der Satz von ASCOLI, eine gleichmäßig konvergente Folge (mit $\alpha = \alpha_n \to 0$) auszusuchen; deren Grenzwert genügt dann dem Gleichungssystem (11), da (12) für $\alpha = \alpha_n \to 0$ offenbar in (11) übergeht. Da man andererseits stetige und beschränkte Funktionen $f(x,y,u,v,w)$ kennt, für welche die charakteristische Anfangswertaufgabe unlösbar ist, werden die Näherungslösungen $u_\alpha, v_\alpha, w_\alpha$ im allgemeinen *nicht* gleichgradig stetig sein. Vielmehr muß man die Klasse der zugelassenen Funktionen $f(x,y,u,v,w)$ weiter einschränken. Wir tun dies zunächst durch die Forderung

$$(14)\quad \begin{cases} |f(x,y,u,v,w) - f(x,y,u,\bar{v},\bar{w})| \leqq g_1(y, |v-\bar{v}|) + g_2(x, |w-\bar{w}|) \\ \qquad \text{für } 0 < x \leqq a,\ 0 < y \leqq b,\ -\infty < u, v, \bar{v}, w, \bar{w} < \infty. \end{cases}$$

Dabei sei $g_1(x,z) \in G'(0,b)$ und $g_2(x,z) \in G'(0,a)$ (s. § 2).

[2a]) Für diesen Ansatz gilt bezüglich des Argumentes u_α das schon oben bei Gl. (10) Gesagte. Dagegen ist es für das Gelingen des Beweises wesentlich, daß bei v_α das Argument $(s, t-\alpha)$ und bei w_α das Argument $(s-\alpha, t)$ steht.

Im Verlauf des nun folgenden Beweises für die gleichgradige Stetigkeit der Funktionen u_α, v_α, w_α wird sich zeigen, daß die Voraussetzung (14) allein nicht ausreicht. Führen wir zur Vereinfachung der nachstehenden Abschätzungen eine Funktion $d(t)$ (Stetigkeitsmodul) ein; für $t \geqq 0$ sei sie stetig und monoton wachsend, und es sei

$$d(0) = 0 \quad \text{und} \quad d(-t) = d(t). \tag{15}$$

Die Funktion $d(t)$ sei so gewählt, daß erstens

$$\begin{cases} |\sigma(x+h) - \sigma(x)| \leqq d(h), & |\sigma'(x+h) - \sigma'(x)| \leqq d(h) \\ |\tau(y+k) - \tau(y)| \leqq d(k), & |\tau'(y+k) - \tau'(y)| \leqq d(k) \end{cases} \tag{16}$$

für alle zugelassenen (d.h. im Definitionsbereich der betreffenden Funktionen liegenden) Werte x, y, $x+h$, $y+k$ ist. Ist K eine Konstante derart, daß

$$|u_\alpha(x,y)| < K, \quad |v_\alpha(x,y)| < K, \quad |w_\alpha(x,y)| < K$$

für alle $\alpha > 0$ und alle zugelassenen (x, y) ist, so gelte zweitens

$$|f(x,y,u,v,w) - f(\bar{x},\bar{y},\bar{u},\bar{v},\bar{w})| \leqq d(x-\bar{x}) + d(y-\bar{y}) + \cdots + d(w-\bar{w}) \tag{17}$$

für $(x, y) \in R$, $(\bar{x}, \bar{y}) \in R$ und $-K < u, \bar{u}, \ldots, \bar{w} < K$. Eine diesen Anforderungen genügende Funktion $d(t)$ kann wegen der gleichmäßigen Stetigkeit von f sicher gefunden werden.

Aus (12) ergeben sich dann die Abschätzungen

$$\begin{cases} |u_\alpha(x+h, y+k) - u_\alpha(x,y)| \leqq d(h) + d(k) + M(a|k| + b|h|) \\ \qquad \text{für } x \leqq a,\ x+h \leqq a,\ 0 \leqq y \leqq b,\ 0 \leqq y+k \leqq b \\ |v_\alpha(x, y+k) - v_\alpha(x,y)| \leqq M|k| \quad \text{für } 0 \leqq x \leqq a,\ y \leqq b,\ y+k \leqq b \\ |w_\alpha(x+h, y) - w_\alpha(x,y)| \leqq M|h| \quad \text{für } x \leqq a,\ x+h \leqq a,\ 0 \leqq y \leqq b; \end{cases} \tag{18}$$

d.h. die Funktionen u_α sind gleichgradig stetig, und ebenso die v_α bezüglich der Veränderlichen y, die w_α bezüglich der Veränderlichen x. Wegen

$$\begin{aligned} |w_\alpha(x+h, y+k) - w_\alpha(x,y)| \leqq\ & |w_\alpha(x+h, y+k) - w_\alpha(x, y+k)| + \\ & + |\tau'(y+k) - \tau'(y)| + |w_\alpha(x, y+k) - w_\alpha(x,y) - \tau'(y+k) + \tau'(y)|, \end{aligned}$$

(16) und der dritten Ungleichung (18) ist die gleichgradige Stetigkeit aller w_α gesichert, wenn das folgende gezeigt werden kann:

Zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$, so daß

$$|w_\alpha(x, y+k) - w_\alpha(x,y) - \tau'(y+k) + \tau'(y)| < \varepsilon \qquad \text{für } |k| < \delta \tag{19}$$

und alle $\alpha > 0$, $-\infty < x \leqq a$, $0 \leqq y \leqq b$, $0 \leqq y+k \leqq b$ ist.

Eine entsprechende Überlegung läßt sich für v_α durchführen; sie führt auf

$$|v_\alpha(x+h, y) - v_\alpha(x,y) - \sigma'(x+h) + \sigma'(x)| < \varepsilon \qquad \text{für } |h| < \delta. \tag{19'}$$

Wir führen nur den Beweis der Behauptung (19) durch, da (19′) ganz entsprechend bewiesen wird. Bezeichnen wir die linke Seite der Ungleichung (19) mit $W_\alpha(x, y, k)$, so ist nach (13)

$$W_\alpha(x,y,k) = |w_\alpha(x, y+k) - w_\alpha(x,y) - \tau'(y+k) + \tau'(y)| = 0 \tag{20}$$

für $x \leqq 0,\ 0 \leqq y \leqq b,\ 0 \leqq y+k \leqq b$

und nach (12)

$$\begin{cases} W_\alpha(x,y,k) \leqq \\ \leqq \int_0^x |f(s, y+k, u_\alpha(s-\alpha, y+k), v_\alpha(s, y+k-\alpha), w_\alpha(s-\alpha, y+k)) - \\ - f(s, y, u_\alpha(s-\alpha, y), v_\alpha(s, y-\alpha), w_\alpha(s-\alpha, y))|\, ds \end{cases} \tag{21}$$

für $0 \leqq x \leqq a,\ 0 \leqq y \leqq b,\ 0 \leqq y+k \leqq b$. Der Integrand in (21) ist

$$\begin{aligned} &\leqq |f(s, y+k, u_\alpha(s-\alpha, y+k), v_\alpha(s, y+k-\alpha), w_\alpha(s-\alpha, y+k)) - \\ &\quad - f(s, y, u_\alpha(s-\alpha, y), v_\alpha(s, y-\alpha), w_\alpha(s-\alpha, y+k))| + \\ &\quad + |f(s, y, u_\alpha(s-\alpha, y), v_\alpha(s, y-\alpha), w_\alpha(s-\alpha, y+k)) - \\ &\quad - f(s, y, u_\alpha(s-\alpha, y), v_\alpha(s, y-\alpha), w_\alpha(s-\alpha, y))|. \end{aligned}$$

Die erste dieser beiden Differenzen ist nach (17) und (18), da f im ersten und letzten Argument übereinstimmt,

$$\begin{aligned} &\leqq d(k) + d(|u_\alpha(s-\alpha, y+k) - u_\alpha(s-\alpha, y)|) + d(|v_\alpha(s, y+k-\alpha) - v_\alpha(s, y-\alpha)|) \\ &\leqq d(k) + d[d(k) + M a |k|] + d(M k), \end{aligned}$$

während die zweite mit Hilfe von (14), (17) und (18) wie folgt abgeschätzt wird (die vier ersten Argumente von f stimmen überein und sind durch Punkte angedeutet):

$$\begin{aligned} &|f(\ldots, w_\alpha(s-\alpha, y+k)) - f(\ldots, w_\alpha(s-\alpha, y))| \\ &\quad \leqq |f(\ldots, w_\alpha(s-\alpha, y+k)) - f(\ldots, w_\alpha(s-\alpha, y+k) + \tau'(y) - \tau'(y+k))| + \\ &\quad + |f(\ldots, w_\alpha(s-\alpha, y+k) + \tau'(y) - \tau'(y+k)) - f(\ldots, w_\alpha(s-\alpha, y))| \\ &\quad \leqq d(d(k)) + g_2(s, |w_\alpha(s-\alpha, y+k) - w_\alpha(s-\alpha, y) + \tau'(y) - \tau'(y+k)|). \end{aligned}$$

Mit der Abkürzung

$$d_1(k) = d(k) + d[d(k) + M a |k|] + d(d(k)) + d(M k)$$

ergibt sich also aus (21)

$$W_\alpha(x,y,k) \leqq d_1(k)\, x + \int_0^x g_2(s, W_\alpha(s-\alpha, y, k))\, ds, \tag{22}$$

gültig für $0 \leqq x \leqq a,\ 0 \leqq y \leqq b,\ 0 \leqq y+k \leqq b$ und jedes $\alpha > 0$. Für jeden positiven Wert von α und δ definieren wir

$$\eta(x) = \eta(x; \alpha, \delta) = \sup\{W_\alpha(x,y,k) \mid 0 \leqq y \leqq b,\ 0 \leqq y+k \leqq b,\ |k| < \delta\};$$

$\eta(x)$ ist definiert für $x \leqq a$ und es ist $\eta = 0$ für $x \leqq 0$. Aus (22) ergibt sich dann wegen der Monotonie von g_2, indem man zuerst auf der rechten, dann auf der linken Seite zum supremum übergeht, die Ungleichung

$$\eta(x) \leqq d_1(\delta)\, x + \int_0^x g_2(s, \eta(s-\alpha))\, ds. \tag{23}$$

Wir machen nun die folgende, die Funktion $g_2(x, z)$ betreffende Voraussetzung A oder genauer $A(0, a)$:

(A) *Ist* $\zeta(x) = \zeta(x; \alpha, \gamma)$ $(\alpha > 0,\ \gamma > 0)$ *eine Lösung der Gleichung*

$$\zeta(x) = \gamma x + \int_0^x g_2(s, \zeta(s-\alpha))\, ds \qquad \text{für } 0 \leqq x \leqq a \tag{24}$$

und $\zeta(x; \alpha, \gamma) = 0$ *für* $x \leqq 0$, *so gelte*[3])

$$\lim_{\gamma \to 0} \zeta(x; \alpha, \gamma) = 0 \qquad \text{gleichmäßig in } x \text{ und } \alpha.$$

Man sieht leicht, daß $\eta(x; \alpha, \delta) < \zeta(x; \alpha, \gamma)$ für $d_1(\delta) < \gamma$ und $0 < x \leqq a$ ist (diese Ungleichung ist nämlich richtig für kleine x, und aus der Annahme $\eta < \zeta$ für $0 < x < \xi$, $\eta(\xi) = \zeta(\xi)$ folgt sofort ein Widerspruch). Infolgedessen ergibt sich aus der Bedingung A, daß $\eta(x; \alpha, \delta)$ und damit auch $W_\alpha(x, y, k)$ gegen Null konvergiert für $\delta \to 0$, und zwar gleichmäßig in x, y, und α, womit (19) bewiesen ist. Das Ergebnis sei im nachstehenden Satz 1 zusammengefaßt.

Satz 1 (Existenzsatz). *Die Funktion* $f(x, y, u, v, w)$ *sei stetig und beschränkt und genüge für* $0 < x \leqq a$, $0 < y \leqq b$, $-\infty < u, v, \bar{v}, w, \bar{w} < \infty$ *der Abschätzung*

$$|f(x, y, u, v, w) - f(x, y, u, \bar{v}, \bar{w})| \leqq g_1(y, |v - \bar{v}|) + g_2(x, |w - \bar{w}|), \tag{14}$$

wobei die Funktion $g_1(x, z) \in G'(0, b)$ *die Voraussetzung* $A(0, b)$ *und die Funktion* $g_2(x, z) \in G'(0, a)$ *die Voraussetzung* $A(0, a)$ *erfüllen möge. Dann besitzt die charakteristische Anfangswertaufgabe für die Differentialgleichung* (1) *mindestens eine Lösung* $u(x, y) \in C^*$.

Wir wollen die etwas undurchsichtige Bedingung A durch andere, leichter anwendbare Existenzbedingungen ersetzen. Betrachten wir die beiden folgenden Aussagen (wobei immer $g(x, z) \in G'(0, a)$ vorausgesetzt ist):

(A_1) *Es sei* $\vartheta(x) = \vartheta(x; \alpha, \gamma)$ *eine Lösung der Integralgleichung*

$$\vartheta(x) = \gamma x + \int_\alpha^x g(s, \vartheta(s))\, ds \qquad \text{für } \alpha \leqq x \leqq a \qquad (\alpha > 0,\ \gamma > 0)$$

oder, was dasselbe bedeutet, eine Lösung der Differentialgleichung

$$\vartheta'(x) = \gamma + g(s, \vartheta) \quad \text{mit} \quad \vartheta(\alpha) = \alpha\gamma \qquad (\alpha \leqq x \leqq a).$$

Dann gelte[4])

$$\lim_{\gamma \to 0} \vartheta(x; \alpha, \gamma) = 0 \qquad \text{gleichmäßig in } x \text{ und } \alpha.$$

[3]) Man überlegt sich leicht, daß ζ für jedes $a > 0$, $\gamma > 0$ existiert.

[4]) Die Existenz von ϑ ist nach dem Satz von Peano gesichert.

(A_2) *Für jedes (hinreichend kleine) positive γ existiere eine im Intervall $0 \leqq x \leqq a$ stetige Funktion $\varphi(x) = \varphi(x;\gamma)$, welche für $0 < x \leqq a$ der Differentialgleichung*

$$\varphi'(x) = \gamma + g(x, \varphi)$$

sowie der Anfangsbedingung $\varphi(0;\gamma) = 0$ genügt, und es gelte

$$\lim_{\gamma \to 0} \varphi(x;\gamma) = 0 \qquad \textit{gleichmäßig in } x.$$

Wir wollen zeigen, daß A in A_1 und A_1 in A_2 enthalten ist. Die Ungleichung

$$\vartheta(x;\alpha,\gamma) < \varphi(x;\bar{\gamma}) \qquad \text{für} \qquad \gamma < \bar{\gamma}$$

ist nämlich richtig in einer rechtsseitigen Umgebung des Punktes $x = \alpha$. Daraus schließt man in bekannter Weise (ähnlich wie oben bei der Ungleichung $\eta < \zeta$) auf ihre Richtigkeit im ganzen Intervall $\alpha \leqq x \leqq a$. Mit ähnlichen Überlegungen zeigt man, daß auch

$$\zeta(x;\alpha,\gamma) < \vartheta(x;\alpha,\bar{\gamma}) \qquad \text{für} \qquad \gamma < \bar{\gamma} \qquad \text{und} \qquad \alpha \leqq x \leqq a$$

ist (man beachte, daß der Integrand in (24) für $0 \leqq x \leqq \alpha$ identisch verschwindet).

Prüfen wir diese Bedingungen an einem einfachen Beispiel

$$g(x,z) = z\left[\frac{\beta}{x} + K(x)\right], \qquad \beta \geqq 0$$

nach ($K(x)$ sei, wie in den Beispielen von § 2, stetig für $0 < x \leqq a$ und summierbar über das Intervall $[0, a]$). Für $\beta < 1$ ist, wie man leicht nachrechnet,

$$\varphi(x;\gamma) = \gamma\, x^{\beta} e^{J(x)} \int_0^x t^{-\beta} e^{-J(t)}\, dt \qquad \text{mit} \qquad J(x) = \int_0^x K(t)\, dt,$$

woraus ersichtlich wird, daß in der Tat A_2 und damit auch der Existenzsatz gilt. Für $\beta = 1$ existiert dagegen, selbst bei Beschränkung auf den Fall $K \equiv 0$, keine Funktion $\varphi(x;\gamma)$ mit den entsprechenden Eigenschaften, und auch A_1 ist, wie man der expliziten Darstellung von $\vartheta(x)$

$$\vartheta(x;\alpha,\gamma) = \gamma\, x\left(1 + \ln\frac{x}{\alpha}\right)$$

entnimmt, nicht gültig. Ob in diesem Fall der Existenzsatz aufrecht erhalten werden kann, bleibt eine offene Frage. Jedoch soll an dieser Stelle schon darauf hingewiesen werden, daß man für $\beta > 1$ *unlösbare* Anfangswertprobleme angeben kann, bei denen f einer Abschätzung (14) mit $g_2(x,z) = \beta z/x$ genügt.

ZUSATZ ZU SATZ 1. *Der Existenzsatz ist gültig, wenn in der Abschätzung* (14) *die Funktionen g_1, g_2 der Bedingung A_1 oder A_2 genügen, insbesondere, wenn sie aus der Klasse $G''(0, b)$ bzw. $G''(0, a)$ sind.*

Wir deuten den nicht schwierigen Beweis der letzten Behauptung kurz an. Es sei also $g(x, z) \in G''$, und es soll gezeigt werden, daß $g(x, z)$ die Eigenschaft

A_2 besitzt. Zu jedem positiven γ gibt es eine in $[0, a]$ stetige Funktion $z(x) = z(x; \gamma)$, welche der Integralgleichung

$$z(x) = \gamma + \int_0^x g\big(t, z(t)\big)\, dt$$

genügt; man wende etwa sukzessive Approximation an, ausgehend von einer (nach § 2, Bedingung $(\dot{\eta})$ sicher existierenden) Funktion $z_0(x)$ mit

$$z_0(x) \geqq \gamma + \int_0^x g\big(t, z_0(t)\big)\, dt.$$

Ebenso erhält man die Funktion $\varphi(x) = \varphi(x; \gamma)$ durch sukzessive Approximation

$$\varphi_{n+1}(x) = \gamma x + \int_0^x g\big(t, \varphi_n(t)\big)\, dt \qquad (n = 0, 1, \ldots).$$

Beginnt man hierbei etwa mit der Funktion $\varphi_0(x) = z(x; \bar{\gamma})$, wobei $\bar{\gamma} > a\gamma$ ist, so folgt

$$\varphi_1(x) = \gamma x + \int_0^x g(t, \varphi_0)\, dt < \bar{\gamma} + \int_0^x g(t, \varphi_0)\, dt = \varphi_0(x),$$

woraus sich in der Tat ergibt, daß die φ_n monoton fallend gegen eine Lösung $\varphi(x; \gamma)$ der Integralgleichung $\varphi(x) = \gamma x + \int_0^x g(t, \varphi)\, dt$ konvergieren. Schließlich ist

$$z(x; \gamma) < z(x; \bar{\gamma}) \quad \text{für} \quad \gamma < \bar{\gamma}$$

und

$$\varphi(x; \gamma) < z(x; \bar{\gamma}) \quad \text{für} \quad a\gamma < \bar{\gamma}$$

sowie

$$\lim_{\gamma \to 0} z(x; \gamma) = 0 \qquad \text{gleichmäßig in } x$$

(die drei letzten Behauptungen lassen sich aus den Integralgleichungen für z und φ leicht herleiten). Hiermit ist gezeigt, daß $g(x, z)$, wie behauptet, die Eigenschaft A_2 hat.

Hiernach ist es also erlaubt, für die Funktionen $g_1(x, z)$, $g_2(x, z)$ die Beispiele (a) und (b) von § 2 zu nehmen. Man erhält so zwei spezielle hinreichende Bedingungen für die Existenz einer Lösung des charakteristischen Anfangswertproblems

$$|f(x, y, u, v, w) - f(x, y, u, \bar{v}, \bar{w})| \leqq K(y)\,|v - \bar{v}| + K(x)\,|w - \bar{w}|$$

(Lipschitz-Bedingung) und

$$|f(x, y, u, v, w) - f(x, y, u, \bar{v}, \bar{w})| \leqq K(y)\,\omega(|v - \bar{v}|) + K(x)\,\omega(|w - \bar{w}|)$$

(Osgood, Montel; vgl. § 2 für die genauen Voraussetzungen).

§ 8. Verallgemeinerungen und Beispiele

Wir wollen uns zunächst von der Voraussetzung, daß f beschränkt ist, frei machen. Man wird dann im allgemeinen, wie einfachste Beispiele zeigen, nicht erwarten dürfen, daß eine Lösung $u(x, y)$ der charakteristischen Anfangs-

wertaufgabe im ganzen Rechteck R existiert. Doch gibt es in diesem Fall eine Lösung wenigstens in einer Umgebung der beiden Charakteristiken $x=0$ und $y=0$, wie man durch Anwendung des folgenden wohlbekannten Kunstgriffs zeigt. Für positives M wird f_M definiert durch

$$f_M(x,y,u,v,w)=\begin{cases} M & \text{für } f>M \\ f(x,y,u,v,w) & \text{für } |f|\leqq M \\ -M & \text{für } f<-M. \end{cases} \tag{25}$$

Mit f genügt auch f_M einer Abschätzung (14). Ist M größer als die Beträge von $\sigma, \sigma', \tau, \tau'$, so ist die Lösung $u(x,y)$ des Anfangswertproblems für die Gleichung

$$u_{xy}=f_M(x,y,u,u_x,u_y)$$

auch eine Lösung des entsprechenden Problems für die Gl. (1), solange u, u_x und u_y dem Betrage nach $\leqq M$ sind, also sicher in einer Umgebung der beiden Anfangscharakteristiken.

Doch kann man in Spezialfällen mehr aussagen. Wir brauchen dazu den folgenden Satz über Differentialungleichungen.

SATZ 2. *Die Funktion $F(x,y,u,v,w)$ sei für $(x,y)\in R$, $u\geqq 0$, $v\geqq 0$, $w\geqq 0$ stetig, monoton wachsend in u, v, w und nichtnegativ, jedoch nicht notwendig beschränkt; es sei also*

$$0\leqq F(x,y,u,v,w)\leqq F(x,y,\bar{u},\bar{v},\bar{w}) \quad \text{für } 0\leqq u\leqq\bar{u},\ 0\leqq v\leqq\bar{v},\ 0\leqq w\leqq\bar{w}. \tag{26}$$

Ferner seien $u(x,y)$, $\bar{u}(x,y)$ zwei Funktionen aus der Klasse C^* mit den Anfangswerten $\sigma(x)$, $\bar{\sigma}(x)$ (für $y=0$) und $\tau(y)$, $\bar{\tau}(y)$ (für $x=0$), und es sei

$$|\sigma(0)|<\bar{\sigma}(0),\ |\sigma'(x)|<\bar{\sigma}'(x) \quad \text{für } 0\leqq x\leqq a,\ |\tau'(y)|<\bar{\tau}'(y) \quad \text{für } 0\leqq y\leqq b. \tag{27}$$

Dann gilt, falls

$$|u_{xy}|\leqq F(x,y,|u|,|u_x|,|u_y|) \quad \text{und} \quad \bar{u}_{xy}\geqq F(x,y,\bar{u},\bar{u}_x,\bar{u}_y) \tag{28}$$

ist, im ganzen Rechteck R

$$|u|<\bar{u},\quad |u_x|<\bar{u}_x,\quad |u_y|<\bar{u}_y. \tag{29}$$

BEWEIS: Ist

$$\xi=\sup\{x_0\,|\,|u|<\bar{u},\ |u_x|<\bar{u}_x,\ |u_y|<\bar{u}_y \quad \text{für} \quad 0\leqq x\leqq x_0,\ 0\leqq y\leqq b\}$$

($\xi=0$, falls kein solches x_0 existiert) und

$$\eta=\sup\{y_0\,|\,|u|<\bar{u},\ |u_x|<\bar{u}_x,\ |u_y|<\bar{u}_y \quad \text{für} \quad x=\xi,\ 0\leqq y\leqq y_0\},$$

so ist nach (27) $(\xi,\eta)\neq(0,0)$ sowie nach (26) und (28)

$$|u_{xy}(x,y)|\leqq F(x,y),|u|,|u_x|,|u_y|)\leqq F(x,y,\bar{u},\bar{u}_x,\bar{u}_y)\leqq\bar{u}_{xy}(x,y)$$

für $0 \leqq x \leqq \xi$, $0 \leqq y \leqq \eta$. Für diese x, y gilt also

$$|u_x(x,y)| = |\sigma'(x) + \int_0^y u_{xy}(x,t)\,dt| < \bar{\sigma}'(x) + \int_0^y \bar{u}_{xy}(x,t)\,dt = \bar{u}_x(x,y),$$

und ähnlich beweist man die beiden anderen Ungleichungen (29). Ist $(\xi, \eta) = (a, b)$, so ist der Satz bewiesen; ist $(\xi, \eta) \neq (a, b)$, so ist ein Widerspruch erreicht, da andererseits nach Definition von ξ und η mindestens eine der Ungleichungen (29) im Punkte (ξ, η) verletzt ist.

Die Bedeutung dieses Satzes für die vorliegende Fragestellung sei an einem einfachen Beispiel erläutert. Die Funktion

$$\bar{u}(x,y) = \varphi(x+y) \quad \text{mit} \quad \varphi(t) = \alpha e^{\alpha t} - 1 \qquad (\alpha > 0)$$

ist, wie eine einfache Rechnung ergibt, eine Lösung der Differentialgleichung

$$\bar{u}_{xy} = \frac{\alpha^2}{1+2\alpha}(1 + \bar{u} + \bar{u}_x + \bar{u}_y) \tag{30}$$

mit den Anfangswerten

$$\bar{\sigma}(x) = \bar{\tau}(x) = \alpha e^{\alpha x} - 1 \geqq \alpha - 1, \quad \bar{\sigma}'(x) = \bar{\tau}'(x) = \alpha^2 e^{\alpha x} \geqq \alpha^2. \tag{31}$$

Der Existenzsatz des vorigen Paragraphen ist deshalb auch gültig, wenn statt der Beschränktheit von f nur gefordert wird, daß

$$|f(x,y,u,v,w)| \leqq A(1 + |u| + |v| + |w|) \tag{32}$$

ist. Wählt man nämlich im obigen Beispiel α so, daß (bei gegebenem Anfangswertproblem)

$$A \leqq \alpha^2/(1+2\alpha), \quad |\sigma(0)| < \alpha - 1, \quad |\sigma'(x)| < \alpha^2, \quad |\tau'(y)| < \alpha^2$$

und M so, daß

$$|\bar{u}| < M, \quad |\bar{u}_x| < M, \quad |\bar{u}_y| < M$$

ist, so kann man den Existenzsatz des vorigen Paragraphen auf die gemäß (25) definierte Funktion f_M anwenden. Es existiert also eine Lösung $u(x, y)$ und aus Satz 2 folgt leicht

$$|u| < M, \quad |u_x| < M, \quad |u_y| < M,$$

d.h. $u(x, y)$ ist auch eine Lösung der ursprünglichen Gl. (1).

Es ist nicht schwer, (32) durch noch allgemeinere Bedingungen zu ersetzen. Wir führen ohne Beweis an:

Die Funktion f genüge einer Abschätzung

$$|f(x,y,u,v,w)| \leqq A(x+y) + B(x+y)h(|u|) + C(x+y)[k(|v|) + k(|w|)]. \tag{32'}$$

Dabei seien die Funktionen $A(t)$, $B(t)$, $C(t)$ stetig und $\geqq 0$ im Intervall $[0, a+b]$, $h(t)$ und $k(t)$ stetig, positiv und monoton wachsend für positive t. Dann gilt der Existenzsatz des vorigen Paragraphen für alle Anfangswerte $\sigma(x)$, $\tau(y)$ mit

$$|\sigma(0)| < C_1, \quad |\sigma'| < C_2, \quad |\tau'| < C_2,$$

falls es eine im ganzen Intervall $[0, a+b]$ *existierende Lösung* $\varphi(t)$ *von*

$$\varphi''(t) = A(t) + B(t)\,h(\varphi) + 2C(t)\,k(\varphi')$$

mit $\varphi(0) = C_1$ *und* $\varphi'(0) = C_2$ *gibt.*

Eine zweite Verallgemeinerung betrifft die Übertragung des Existenzsatzes auf Systeme von Differentialgleichungen in dem in Teil I, § 5 näher präzisierten Sinne. Diese Übertragung ist in demselben Umfang wie bei den Eindeutigkeitssätzen (also bei Zugrundelegung einer beliebigen Norm; vgl. § 5) möglich. Die Beweise gelten ungeändert auch für diesen Fall, wenn man an den entsprechenden Stellen die Absolutstriche durch Doppelstriche ersetzt.

Wir schließen diesen Paragraphen mit einigen Beispielen ab. Betrachten wir Funktionen $f(x, y, u, v, w)$, die nur von x, y, w abhängen und in y im engeren Sinne monoton wachsend sind:

$$(33)\quad f(x, y, u, v, w) = h(x, y, w) \quad \text{mit} \quad h(x, y, w) < h(x, \bar{y}, w) \quad \text{für } y < \bar{y}.$$

Die charakteristische Anfangswertaufgabe ist dann nach Satz 1 sicher lösbar, wenn h stetig und beschränkt sowie

$$(34)\qquad |h(x, y, w) - h(x, y, \bar{w})| \leqq g_2(x, |w - \bar{w}|)$$

für alle $(x, y) \in R$ und beliebige $w, \bar{w}$ ist.

Betrachten wir nun die *gewöhnliche* Differentialgleichung für eine Funktion $w(x)$

$$(35)\qquad w'(x) = h(x, y, w),$$

in der y als Parameter aufzufassen ist. Diese Gleichung hat für jeden Anfangswert $w(0)$ eine in $[0, a]$ existierende Lösung $w(x)$, und die Bedingung (34) ist gerade die allgemeine Bedingung von KAMKE [7, S. 139] (nicht in voller Allgemeinheit, da nicht nur $g_2 \in G$, sondern $g_2 \in G'$ und (A) vorausgesetzt ist) für die Eindeutigkeit dieser Lösung. Das ist kein zufälliger, durch unsere spezielle Beweismethode hervorgerufener Sachverhalt; *vielmehr besteht eine enge Beziehung zwischen dem Existenzproblem für die charakteristische Anfangswertaufgabe der hyperbolischen Differentialgleichung*

$$(36)\qquad u_{xy} = h(x, y, u_y)$$

und dem Eindeutigkeitsproblem für die gewöhnliche Differentialgleichung (35). Wir behaupten:

Wenn es ein $y = \eta$ $(0 < \eta < b)$ *und eine Zahl* α *gibt, so daß die Differentialgleichung* (35) *mehr als eine Lösung* $w(x)$ *mit dem Anfangswert* $w(0) = \alpha$ *besitzt, dann ist die charakteristische Anfangswertaufgabe für die Gl.* (36) *im allgemeinen nicht lösbar.*

Genauer gesagt, besitzt sie sicher keine Lösung, wenn $\tau'(\eta) = \alpha$ und τ' monoton wachsend in einer Umgebung $|y - \eta| < \delta$ des Punktes η ist. Nehmen wir, im Gegensatz zu dieser Behauptung, an, die Funktion $u(x, y)$ sei unter den eben formulierten Voraussetzungen bezüglich $\tau(y)$ eine Lösung des

Anfangswertproblems der Gl. (36). Das Maximal- bzw. Minimalintegral von (35) werde mit $W(x)$ bzw. $w(x)$ bezeichnet, $W(0) = w(0) = \alpha$. Dann hat für jedes feste y mit $\eta < y < \eta + \delta$ die Funktion $Z(x) = u_y(x, y)$ die Eigenschaften

$$Z'(x) = h(x, y, Z) > h(x, \eta, Z) \quad \text{und} \quad Z(0) = \tau'(y) \geqq \alpha;$$

entsprechend gilt, wenn $\eta - \delta < y < \eta$ und $z(x) = u_y(x, y)$ ist,

$$z'(x) = h(x, y, z) < h(x, \eta, z) \quad \text{und} \quad z(0) = \tau'(y) \leqq \alpha.$$

Es ist also

$$W' = h(x, y, W), \quad Z' > h(x, y, Z) \quad \text{und} \quad Z(0) \geqq W(0).$$

Daraus schließt man in bekannter Weise, daß $W(x) < Z(x)$ für $0 < x \leqq a$ ist, und ebenso erhält man die Ungleichung $z(x) < w(x)$ für $0 < x \leqq a$. Insgesamt hat sich also ergeben, daß

$$u_y(x, y) \geqq W(x) \qquad \text{für } y > \eta$$

und

$$u_y(x, y) \leqq w(x) \qquad \text{für } y < \eta$$

ist. Diese Ungleichungen sind mit der Stetigkeit von u_y unverträglich, woraus die Behauptung folgt.

So ist z. B. das Anfangswertproblem

$$u_{xy} = \sqrt{|u_y|} + y - \eta, \qquad u(0, y) = (y - \eta)^2, \qquad u(x, 0) = \sigma(x) \text{ beliebig}$$

nicht lösbar ($0 < \eta < b$). In diesem Fall ist $w(x) = 0$, $W(x) = x^2/4$. Allgemeiner kann man aus jedem *mehrdeutigen* Anfangswertproblem

$$w'(x) = f(x, w), \qquad w(0) = \alpha$$

unendlich viele *unlösbare* Anfangswertprobleme für die hyperbolische Differentialgleichung

$$u_{xy} = f(x, u_y) + k(y)$$

($k(t)$ monoton wachsend und $k(\eta) = 0$) ableiten.

§ 9. Das Picardsche Iterationsverfahren

Picard hat in seiner klassischen Arbeit [1] gezeigt, daß man eine Lösung der Integralgleichung (11) und damit auch der charakteristischen Anfangswertaufgabe als Limes der durch

$$(37) \quad u_0(x, y) = \sigma(x) + \tau(y) - \sigma(0), \qquad v_0(x, y) = \sigma'(x), \qquad w_0(x, y) = \tau'(y)$$

und

$$(38) \quad \begin{cases} u_{n+1}(x, y) = \sigma(x) + \tau(y) - \sigma(0) + \int_0^x \int_0^y f(s, t, u_n, v_n, w_n)\, ds\, dt \\ v_{n+1}(x, y) = \sigma'(x) + \int_0^y f(x, t, u_n, v_n, w_n)\, dt \\ w_{n+1}(x, y) = \tau'(y) + \int_0^x f(s, y, u_n, v_n, w_n)\, ds \end{cases}$$

($n = 0, 1, 2, \ldots$) definierten und gleichmäßig konvergenten Funktionenfolge (u_n, v_n, w_n) erhalten kann, falls $f(x, y, u, v, w)$ in u, v und w einer Lipschitz-Bedingung genügt[5]). Das Problem, welches wir in diesem Paragraphen untersuchen wollen, lautet: Unter welchen allgemeineren Bedingungen konvergiert die durch (38) definierte (wir sagen auch, durch sukzessive Approximation gewonnene) Funktionenfolge gegen eine Lösung der Integralgleichung (11) (wobei wir statt (37) gelegentlich auch eine andere erste Approximation benützen)?

Beweistechnisch ist das Problem damit noch nicht eindeutig gestellt. Man kann nämlich erstens annehmen, es sei (u, v, w) eine Lösung von (11), und dann nach Bedingungen suchen, unter denen

$$D_n(x, y) = |u - u_n| + |v - v_n| + |w - w_n| \to 0 \qquad (n \to \infty) \tag{39}$$

strebt. In diesem Sinne hat NAGUMO [6] das entsprechende Problem bei der gewöhnlichen Differentialgleichung (2) untersucht. Man kann sich zweitens die Aufgabe stellen, *ohne Benützung einer Lösung* nachzuweisen, daß die Folge (u_n, v_n, w_n) dem Cauchyschen Konvergenzkriterium genügt und gegen eine Lösung konvergiert; diese Aufgabe ist im allgemeinen schwieriger und enthält zugleich einen Existenzsatz. Die üblichen Beweise für die Konvergenz der sukzessiven Approximation (KAMKE [7, S. 403f.], COURANT-HILBERT [13, S. 319f.]) gehören zur zweiten Art.

Was den Konvergenzbeweis erster Art anbetrifft, so ist leicht zu sehen, daß er unter jeder der in Teil I angegebenen Eindeutigkeitsbedingungen durchführbar ist. Wir skizzieren den Beweis am Beispiel von Satz I.7. Es genüge also f der Abschätzung (I.28). Dann wird behauptet, daß entsprechend zur Formel (I.30)

$$D_n(x, y) \leqq \varphi_n(x + y)$$

ist (D_n ist in (39), φ_n im Beweis zu Satz I.7 definiert). Wie bei (I.29) wählen wir φ_0 so, daß die Behauptung für $n = 0$ richtig ist. Der Induktionsbeweis ist wörtlich wie in Satz I.7 wenn man in den entsprechenden Formeln $D(x, y)$ durch $D_n(x, y)$, $\delta u(x, y)$ durch $|u_n(x, y) - u(x, y)|$, $\delta u_x(x, y)$ durch $|v_n(x, y) - v(x, y)|, \ldots$ ersetzt.

Ebenso ist der Beweis von Satz I.4 ohne weiteres übertragbar. Bei Satz I.5 und Satz I.6 tritt eine Schwierigkeit auf. Es wurde bei den Beweisen zu diesen beiden Sätzen zunächst festgestellt, daß, wenn u und $\bar{u}$ zwei Lösungen sind, $u = \bar{u}$, $u_x = \bar{u}_x$ und $u_y = \bar{u}_y$ auf den Charakteristiken $x = 0$ und $y = 0$ gilt. Entsprechend benötigen wir hier das folgende. Ist (u, v, w) eine Lösung von (11), so ist

$$u = u_0, \quad v = v_0, \quad w = w_0 \quad \text{für } x = 0 \text{ und für } y = 0. \tag{40}$$

---

[5]) PICARD führt a. a. O. S. 359 zuerst das allgemeine Problem auf den Fall $\sigma = \tau = 0$ zurück und beginnt dann mit $u_0 = v_0 = w_0 = 0$ (in unserer Schreibweise).

Dies ist aber bei der in (37) getroffenen Wahl im allgemeinen nicht richtig. Man muß daher so vorgehen. Die Funktionen

$$v(0, y) = h(y) \quad \text{und} \quad w(x, 0) = k(x)$$

sind (s. (11) oder den Beweis von Satz I.4) als Lösungen einer *gewöhnlichen* Differentialgleichung ohne Kenntnis von u, v, w eindeutig bestimmbar[6]). Setzt man mit ihrer Hilfe

$$(41) \quad \begin{cases} u_0(x, y) = \sigma(x) + \tau(y) - \sigma(0) + x\,[h(y) - h(0)] + y\,[k(x) - k(0)] - x\,y\,h'(0) \\ v_0(x, y) = \sigma'(x) + h(y) - h(0) + y\,k'(x) - y\,h'(0) \\ w_0(x, y) = \tau'(y) + x\,h'(y) + k(x) - k(0) - x\,h'(0), \end{cases}$$

so ist der Bedingung (40) Genüge getan, und die Beweise von Satz I.5 und, Satz I.6 lassen sich ohne Schwierigkeit übersetzen (man beachte $\sigma'(0) = h(0)$ $\tau'(0) = k(0)$, $h'(0) = k'(0)$). Fassen wir die Ergebnisse im folgenden Satz zusammen.

Satz 3. *Das charakteristische Anfangswertproblem für die Gl.* (1) *besitze eine Lösung* $u(x, y)$, *und die Funktion* $f(x, y, u, v, w)$ *genüge einer der Voraussetzungen von Satz I.4 bis Satz I.7. Dann konvergieren die durch sukzessive Approximation gewonnenen Funktionenfolgen* u_n, v_n, w_n *gleichmäßig in* R *gegen* $u(x, y)$ *bzw.* $u_x(x, y)$ *bzw.* $u_y(x, y)$, *wenn man von einer geeigneten ersten Approximation*[6]) *— etwa gemäß* (41) *— ausgeht.*

Diese vorsichtige Formulierung von Satz 3 ist notwendig, weil der Existenzsatz von § 7 nicht in allen Fällen anwendbar ist. Setzt man etwa in Satz I.5 $\alpha = \beta = 0$, $\gamma = 1$, so ist, wie in § 7 näher ausgeführt wurde, nicht bekannt, ob eine Lösung existiert. Satz 3 sagt in diesem Fall lediglich aus: Wenn eine Lösung vorhanden ist, dann konvergieren die sukzessiven Approximationen gleichmäßig gegen diese Lösung. Satz 3 enthält, grob gesprochen, eine Übertragung der Ergebnisse von Nagumo [6] von gewöhnlichen Differentialgleichungen auf das vorliegende hyperbolische Problem.

Es soll wenigstens skizziert werden, und zwar wiederum am Beispiel des Satzes I.7, wie ein Konvergenzbeweis zweiter Art, also ein Existenzbeweis durch sukzessive Approximation, durchgeführt werden kann. Die Funktion f genüge also einer Abschätzung (I.28), und die sukzessiven Approximationen seien gemäß (37), (38) definiert. Dann gilt für jede natürliche Zahl k, wenn M eine Schranke für den Betrag von f ist,

$$(42) \qquad |u_k - u_0| \leqq M\,x\,y, \qquad |v_k - v_0| \leqq M\,y, \qquad |w_k - w_0| \leqq M\,x.$$

Wir wollen zeigen, daß für die Funktion

$$D_{kn}(x, y) = |u_{n+k} - u_n| + |v_{n+k} - v_n| + |w_{n+k} - w_n|$$

[6]) Hierbei wird also die Integration einer gewöhnlichen Differentialgleichung als ein einfacheres Problem betrachtet, dessen Lösung als bekannt vorausgesetzt werden darf. Übrigens sind die Voraussetzungen derart, daß man die Funktionen $h(y)$ und $k(x)$ durch sukzessive Approximation gewinnen kann.

gleichmäßig in k eine Abschätzung

$$D_{kn}(x,y) \leqq \varphi_{n-1}(x+y) \tag{43}$$

gilt, wobei (mit den Bezeichnungen vom Beweis zu Satz I.7)

$$\varphi_0(t) \in \Phi \quad \text{und} \quad \varphi_n(t) = K_1^n \varphi_0$$

ist. Die durch

$$\psi(t) = \sup\{|f(x,y,u_0(x,y),v_0(x,y),w_0(x,y)) - f(x,y,u,v,w)|;\ (x,y)\in R,$$
$$0 \leqq x+y \leqq t,\ |u-u_0(x,y)| \leqq M\,x\,y,\ |v-v_0(x,y)| \leqq M\,y,$$
$$|w-w_0(x,y)| \leqq M\,x\}$$

definierte Funktion $\psi(t)$ ist stetig in $[0, a+b]$, und es ist $\psi(0)=0$. Also ist

$$|f(x,y,u_k,v_k,w_k) - f(x,y,u_0,v_0,w_0)| \leqq \psi(x+y).$$

Durch Integration dieser Ungleichung erkennt man, daß eine von k unabhängige Abschätzung

$$|u_{k+1}-u_1| + |v_{k+1}-v_1| + |w_{k+1}-w_1| \leqq \varphi_0(x+y) \qquad (\varphi_0 \in \Phi),$$

d.h. die Abschätzung (43) für $n=1$, möglich ist. Der Induktionsbeweis für die behauptete Ungleichung (43) läßt sich dann wie bei Satz I.7 durchführen. Es ergibt sich daraus, daß die Folge der sukzessiven Approximationen dem Cauchyschen Konvergenzkriterium genügt, also gleichmäßig konvergent ist. Der Nachweis, daß der Grenzwert eine Lösung der Gl. (11) darstellt, bereitet keine Schwierigkeit.

Wir schließen mit einigen Bemerkungen zum vorstehenden Konvergenzbeweis. Bei den bekannten Konvergenzbeweisen wird vorausgesetzt, daß f in den Variablen u, v, w gleichmäßig einer Lipschitz-Bedingung genügt. Es wird dann nicht nur gezeigt, daß die u_n, v_n, w_n gleichmäßig konvergieren, sondern, daß sogar die Reihe

$$\sum_{n=0}^{\infty} \{|u_{n+1}-u_n| + |v_{n+1}-v_n| + |w_{n+1}-w_n|\}$$

konvergiert. Das ist aber für den Konvergenzbeweis gar nicht notwendig und unter den allgemeineren Voraussetzungen von Satz I.7 möglicherweise gar nicht richtig. Bei der obigen Beweisanordnung kommt es — und das ist die Hauptschwierigkeit — darauf an, eine für alle k gültige erste Abschätzung von $|u_k-u_0|$ oder $|u_{k+1}-u_1|$ (und ebenso für v_k und w_k) herzuleiten. Diese erste Abschätzung tritt bei der Übertragung der Beweise von Satz I.4 bis Satz I.7 an die Stelle der ersten Abschätzung der Differenz zweier Lösungen $|u-\bar{u}|$. Wir haben oben wenigstens angedeutet, wie man im Falle des Satzes I.7 diese erste Abschätzung, die die Stelle von (I.29) einnimmt, erhalten kann. Bei Satz I.5 und I.6 wird jedoch eine schärfere erste Abschätzung (I.21) verlangt, und wir wissen nicht, ob eine entsprechende Abschätzung von $|u_k-u_0|, \ldots$ gleichmäßig in k möglich ist. Macht man jedoch schärfere Voraussetzungen, etwa

bei Satz I.5: $\alpha+\beta+\gamma<1$ *statt* $\alpha+\beta+\gamma\leqq 1$,

bei Satz I.6: $g(x,z)\in G''(0,c)$ *statt* $g(x,z)\in G'(0,c)$,

so genügt die durch (42) gegebene erste Abschätzung, und die Beweise sind übertragbar. Unter diesen Bedingungen kann man also die Existenz einer Lösung durch sukzessive Approximation nachweisen. Man erhält jedoch auf diese Weise keine neuen, über die Ergebnisse von § 7 hinausgehenden Existenzaussagen.

Literaturverzeichnis (Fortsetzung von Teil I)

[*11*] LEEHEY, P.: On the existence of not necessarily unique solutions of classical hyperbolic boundary value problems for non-linear partial differential equations in two independent variables. Ph. D. thesis, Brown University, June 1950. — [*12*] TONELLI, L.: Sulle equazioni integrali di Volterra. Mem. Accad. delle Sci. Bologna, Cl. Sci. Fis., Ser. 8 **5**, 59—64 (1927/28). — [*13*] COURANT, R., u. D. HILBERT: Methoden der mathematischen Physik, Bd. 2. Berlin 1937. — [*14*] ROSENBLATT, A.: Über die Existenz von Integralen gewöhnlicher Differentialgleichungen. Ark. Mat., Astronom. Fys. **5**, n. 2 (1909). — [*15*] CONTI, R.: Sul problema di Darboux per l'equazione $z_{xy}=f(x, y, z, z_x, z_y)$. Ann. Univ. Ferrara (Nuova ser.), Ser. 7, Sc. Mat. **2**, 129—140 (1953).

Institute for Fluid Dynamics and Applied Mathematics University of Maryland, College Park/Md. (USA) und

Institut für Angewandte Mathematik der Technischen Hochschule Karlsruhe

(Eingegangen am 9. Februar 1959)

BAER, R.
Math. Zeitschr. 71, 454–457 (1959)

Direkte Produkte von Gruppen teilerfremder Ordnung

Von
REINHOLD BAER

FROBENIUS hat gezeigt, daß eine Gruppe der Ordnung hk mit teilerfremden h und k dann und nur dann das direkte Produkt einer Gruppe der Ordnung h und einer der Ordnung k ist, wenn sie genau h Lösungen der Gleichung $x^h = 1$ und genau k Lösungen der Gleichung $x^k = 1$ enthält. Wir wollen im folgenden diesem Kriterium ein anderes zur Seite stellen. Um dieses in den rechten Zusammenhang einzuordnen, es einfach auszusprechen und seinen Beweis übersichtlich zu führen, wird es bequem sein, an einige Begriffsbildungen zu erinnern.

Ist $\mathfrak{r}$ eine Menge von Primzahlen, so nennen wir das Gruppenelement g ein $\mathfrak{r}$-Element, wenn seine Ordnung $o(g)$ nur durch Primzahlen aus $\mathfrak{r}$ teilbar ist; und entsprechend heiße die Gruppe G eine $\mathfrak{r}$-Gruppe, wenn alle Primteiler ihrer Ordnung $o(G)$ in $\mathfrak{r}$ liegen, eine Eigenschaft die damit äquivalent ist, daß G nur aus $\mathfrak{r}$-Elementen besteht. Allgemeiner verstehen wir unter einer $\mathfrak{r}$-abgeschlossenen Gruppe eine Gruppe G derart, daß Produkte von $\mathfrak{r}$-Elementen wieder $\mathfrak{r}$-Elemente sind; und dies ist gleichwertig damit, daß die Gesamtheit $G_\mathfrak{r}$ der $\mathfrak{r}$-Elemente aus G eine [charakteristische] $\mathfrak{r}$-Untergruppe von G ist [deren Index $[G:G_\mathfrak{r}]$ dann natürlich zu jeder Primzahl aus $\mathfrak{r}$ teilerfremd ist].

Verstehen wir unter $\mathfrak{r}'$ die zu $\mathfrak{r}$ komplementäre Primzahlmenge, so ist G offenbar dann und nur dann gleichzeitig $\mathfrak{r}$- und $\mathfrak{r}'$-abgeschlossen, wenn G das direkte Produkt von $G_\mathfrak{r}$ und $G_{\mathfrak{r}'}$ ist; und damit haben wir das Zerlegungsproblem in das Abgeschlossenheitsproblem eingeordnet.

Ist U eine Untergruppe der Gruppe G, so sei $\mathfrak{N}U$ ihr Normalisator und $\mathfrak{C}U$ ihr Zentralisator in G. Die Faktorgruppe $\mathfrak{N}U/\mathfrak{C}U$ ist dann im wesentlichen die Gruppe der von G in seiner Untergruppe U induzierten Automorphismen. Die Gruppe G wollen wir nun $\mathfrak{r}$-homogen nennen, wenn $\mathfrak{N}U/\mathfrak{C}U$ für jede $\mathfrak{r}$-Untergruppe U von G eine $\mathfrak{r}$-Gruppe ist. Es ist leicht einzusehen, daß $\mathfrak{r}$-abgeschlossene Gruppen notwendig auch $\mathfrak{r}'$-homogen sind; s. etwa BAER [*1*; Lemma 2.1]. Ist also die Gruppe G direktes Produkt einer $\mathfrak{r}$-Gruppe und einer $\mathfrak{r}'$-Gruppe, so ist ja G auch $\mathfrak{r}$- und $\mathfrak{r}'$-abgeschlossen; und hieraus folgt dann, daß G auch $\mathfrak{r}'$- und $\mathfrak{r}$-homogen ist. Ob die Umkehrung gilt [oder nicht], haben wir nicht entscheiden können; aber wir wollen wenigstens das folgende Kriterium beweisen.

SATZ. *Die Gruppe G ist dann und nur dann das direkte Produkt einer $\mathfrak{r}$-Gruppe und einer $\mathfrak{r}'$-Gruppe, wenn*

(a) *G sowohl $\mathfrak{r}$- wie auch $\mathfrak{r}'$-homogen ist und*

(b) ***eine einfache Faktorgruppe einer Untergruppe von G nicht gleichzeitig eine nicht-nilpotente $\mathfrak{r}$-Untergruppe und eine nicht-nilpotente $\mathfrak{r}'$-Untergruppe besitzen kann.***

Auf die Notwendigkeit der Bedingung (a) haben wir schon hingewiesen, während (b) sich trivialerweise daraus ergibt, daß eine derartige Faktorgruppe entweder eine $\mathfrak{r}$-Gruppe oder eine $\mathfrak{r}'$-Gruppe ist, wenn G gleichzeitig $\mathfrak{r}$- und $\mathfrak{r}'$-abgeschlossen ist, da sich die Abgeschlossenheitseigenschaft auf Unter- und Faktorgruppen vererbt.

LEMMA. *Ist die einfache Gruppe E weder eine $\mathfrak{r}$- noch eine $\mathfrak{r}'$-Gruppe, ist jede echte Untergruppe von E entweder eine $\mathfrak{r}$- oder eine $\mathfrak{r}'$-Gruppe, so gibt es nicht-nilpotente $\mathfrak{r}$-Untergruppen und nicht-nilpotente $\mathfrak{r}'$-Untergruppen von E.*

BEWEIS. Aus der ersten Voraussetzung folgt insbesondere, daß E weder $\mathfrak{r}$- noch $\mathfrak{r}'$-abgeschlossen sein kann. Ist $U \neq 1$ eine $\mathfrak{r}$-Untergruppe von E, so ist U kein Normalteiler von E. Also ist $1 < U \leqq \mathfrak{N}U < E$; und aus der zweiten Voraussetzung folgt, daß auch $\mathfrak{N}U$ eine $\mathfrak{r}$-Gruppe ist. Hieraus folgt insbesondere die $\mathfrak{r}$-Homogenität von E; und die $\mathfrak{r}'$-Homogenität folgt aus Symmetriegründen. Eine $\mathfrak{r}$-homogene Gruppe, deren sämtliche $\mathfrak{r}$-Untergruppen nilpotent sind, ist aber $\mathfrak{r}'$-abgeschlossen; s. BAER [2; Satz A]. Da E nicht $\mathfrak{r}'$-abgeschlossen ist, so gibt es $\mathfrak{r}$-Untergruppen von E, die nicht nilpotent sind; und aus Symmetriegründen folgt die Existenz von $\mathfrak{r}'$-Untergruppen, die nicht nilpotent sind.

BEMERKUNG. Es scheint eine offene Frage zu sein, ob es überhaupt eine einfache Gruppe E gibt, die den Voraussetzungen unseres Lemma genügt. Aus diesem Lemma folgt z. B. die folgende Eigenschaft einer derartigen einfachen Gruppe E: Mindestens zwei Primteiler von $o(E)$ liegen in $\mathfrak{r}$ und mindestens zwei Primteiler von $o(E)$ liegen in $\mathfrak{r}'$.

Aus dem Lemma ergibt sich, daß die Bedingung (b) des Satzes die Gültigkeit der folgenden Bedingung nach sich zieht:

(b') *Ist die einfache Gruppe E Faktorgruppe einer Untergruppe von G, ist jede echte Untergruppe von E entweder eine $\mathfrak{r}$-Gruppe oder eine $\mathfrak{r}'$-Gruppe, so ist E selbst entweder eine $\mathfrak{r}$-Gruppe oder eine $\mathfrak{r}'$-Gruppe.*

Es ist unbekannt, ob (b') allgemein gilt oder ob wenigstens (b') eine Folge von (a) ist. — Aus dem Lemma ergibt sich jedenfalls, daß es zum Beweis unseres Satzes genügt nachzuweisen, daß Gruppen mit den Eigenschaften (a) und (b') direkte Produkte von $\mathfrak{r}$- und $\mathfrak{r}'$-Gruppen sind; und dem Beweis dieser Tatsache wollen wir uns jetzt zuwenden.

Wäre dies falsch, so existierte eine Gruppe K kleinster Ordnung mit den folgenden Eigenschaften:

(1) K ist sowohl $\mathfrak{r}$-homogen wie auch $\mathfrak{r}'$-homogen.

(2) Ist die einfache Gruppe E Faktorgruppe einer Untergruppe von K, ist jede echte Untergruppe von E entweder eine $\mathfrak{r}$-Gruppe oder eine $\mathfrak{r}'$-Gruppe, so ist E selbst entweder eine $\mathfrak{r}$-Gruppe oder eine $\mathfrak{r}'$-Gruppe.

(3) K ist nicht das direkte Produkt einer $\mathfrak{r}$-Gruppe und einer $\mathfrak{r}'$-Gruppe.

Untergruppen und Faktorgruppen $\mathfrak{x}$-homogener Gruppen sind $\mathfrak{x}$-homogen; s. BAER [*1*; Lemma 2.3]. Also vererbt sich Bedingung (1) auf Untergruppen und Faktorgruppen. Bedingung (2) vererbt sich offenbar ebenfalls auf Untergruppen und Faktorgruppen. Also folgt aus der Minimalität von K:

(4) Jede echte Untergruppe und jede echte Faktorgruppe von K ist das direkte Produkt einer $\mathfrak{r}$-Gruppe und einer $\mathfrak{r}'$-Gruppe.

Wäre K nicht einfach, so existierte ein $1<J<K$ erfüllender Normalteiler J von K. Aus (4) folgt, daß sowohl J wie auch K/J direkte Produkte von $\mathfrak{r}$- und $\mathfrak{r}'$-Gruppen sind. Insbesondere ist also die nach (1) $\mathfrak{r}'$-homogene Gruppe K Erweiterung der $\mathfrak{r}$-abgeschlossenen Gruppe J durch die $\mathfrak{r}$-abgeschlossene Gruppe K/J; und hieraus folgt nach BAER [*1*; Lemma 2.4] die $\mathfrak{r}$-Abgeschlossenheit von K. Ebenso erschließt man die $\mathfrak{r}'$-Abgeschlossenheit von K; und diese beiden Abgeschlossenheitseigenschaften zusammen sind äquivalent damit, daß K das direkte Produkt einer $\mathfrak{r}$- und einer $\mathfrak{r}'$-Gruppe ist. Dies widerspricht (3); und damit haben wir bewiesen:

(5) K ist einfach.

Ist U eine echte Untergruppe von K, so ist U wegen (4) das direkte Produkt einer $\mathfrak{r}$-Untergruppe U^* und einer $\mathfrak{r}'$-Untergruppe U^{**}; und sowohl die $\mathfrak{r}$-Komponente U^* wie auch die $\mathfrak{r}'$-Komponente U^{**} ist durch U eindeutig bestimmt. Ist sowohl $U^*\neq 1$ wie auch $U^{**}\neq 1$, so wollen wir die echte Untergruppe U von K eine gemischte Untergruppe von K nennen. Gibt es überhaupt gemischte Untergruppe von K, so gibt es auch gemischte Untergruppen, die gleichzeitig maximale Untergruppen von K sind; und unter diesen gemischten maximalen Untergruppen von K gibt es eine W, deren $\mathfrak{r}$-Komponente W^* maximale Ordnung hat.

Da W als gemischte Untergruppe eine echte Untergruppe von K ist, so ist W wegen (5) kein Normalteiler von K; und wir können $W=\mathfrak{N}W$ aus der Maximalität von W erschließen.

Sei nun x ein nicht in W enthaltenes Element aus K. Dann ist $W\neq W^x$. Angenommen nun es wäre $W^{**}\cap W^{**x}\neq 1$. Dann gäbe es ein Element $t\neq 1$ in $W^{**}\cap W^{**x}$. Da t in W^{**} liegt und $W=W^*\otimes W^{**}$ ist, so ist W^* im Zentralisator $T=\mathfrak{C}t$ von t enthalten; und da t in W^{**x} liegt und $W^x=W^{*x}\otimes W^{**x}$ ist, so ist auch W^{*x} in $T=\mathfrak{C}t$ enthalten. Da $t\neq 1$ im Zentrum von T liegt, so ist $T\neq 1$ und $T\neq K$, da ja K einfach und wegen (3) auch nicht abelsch ist. Als echte Untergruppe von K liegt T in einer maximalen Untergruppe S von K. Da sowohl das $\mathfrak{r}'$-Element $t\neq 1$ wie auch die $\mathfrak{r}$-Untergruppe $W^*\neq 1$ [und $W^{*x}\neq 1$] in T und S enthalten sind, so ist S eine gemischte maximale Untergruppe von K. Es gilt $\{W^*, W^{*x}\}\leqq S^*$ und $o(S^*)\leqq o(W^*)$ wegen der Maximalität von W^*. Folglich ist $S^*=W^*=W^{*x}$. Da S^* ein direkter Faktor von S ist, so liegt S im Normalisator von S^*. Da S^* eine echte Untergruppe von K ist, da K wegen (5) einfach ist, und da schließlich S eine maximale Untergruppe von K ist, so wird $S=\mathfrak{N}S^*$. Nun enthält aber $\mathfrak{N}S^*=\mathfrak{N}W^*=\mathfrak{N}(W^{*x})$ sicherlich auch W^{**} und W^{**x}; und daraus folgt, daß $W=W^*\otimes W^{**}$ und $W^x=W^{*x}\otimes W^{**x}$ beide in $S=\mathfrak{N}S^*$ liegen. Aus der

Maximalität der Untergruppen S, W, W^x folgt schließlich $W = S = W^x$; und dies ist ein Widerspruch. Damit haben wir $W^{**} \cap W^{**x} = 1$ gezeigt. Ist g ein Element aus $W \cap W^x$, so ist $g = g^* g^{**}$, wo g^* ein $\mathfrak{r}$-Element und g^{**} ein $\mathfrak{r}'$-Element ist, und wo g^* und g^{**} beide in $\{g\}$ liegen. Dann liegt aber g^{**} in $W^{**} \cap W^{**x} = 1$. Also ist $g^{**} = 1$ und $g = g^*$ ist ein $\mathfrak{r}$-Element. Damit haben wir gezeigt, daß $W \cap W^x \leqq W^*$ für jedes nicht in W gelegene x aus K gilt. Da W^* ein Normalteiler von W ist, so können wir den Satz von WIELANDT-FROBENIUS anwenden; s. WIELANDT, p. 276, Satz 1. Also existiert ein Normalteiler L von K mit folgenden Eigenschaften: $K = LW$, $W^* = L \cap W$. Aus $1 < W^* < W$ folgt $1 < L < K$; und dies widerspricht der Einfachheit von K [s. (5)]. Damit haben wir aber die Nichtexistenz gemischter [echter] Untergruppen von K dargetan. Anwendung von (b') ergibt nun, daß K entweder eine $\mathfrak{r}$-Gruppe oder eine $\mathfrak{r}'$-Gruppe ist; und dies widerspricht (3). Damit haben wir gezeigt, daß jede Gruppe mit den Eigenschaften (a) und (b') direktes Produkt einer $\mathfrak{r}$-Gruppe und einer $\mathfrak{r}'$-Gruppe ist, q.e.d.

BEMERKUNG. Aus dem Beweis geht die Gültigkeit der folgenden Tatsache hervor:

Sind alle echten Untergruppen der einfachen Gruppe K direkte Produkte von $\mathfrak{r}$- und $\mathfrak{r}'$-Gruppen, so ist jede echte Untergruppe von K entweder eine $\mathfrak{r}$-Gruppe oder eine $\mathfrak{r}'$-Gruppe.

Literatur

BAER, R.: [1] Closure and Dispersion of finite groups. Illinois J. Math. **2**, 619–640 (1959). – [2] Kriterien für die Abgeschlossenheit endlicher Gruppen. Math. Z. **71**, 325–334 (1959). – FROBENIUS, G.: Über auflösbare Gruppen. IV. Sitzgsber. preuß. Akad. Wiss. 1216–1230 (1901). – WIELANDT, H.: Über die Existenz von Normalteilern in endlichen Gruppen. Math. Nachr. **18**, 274–280 (1958).

Frankfurt a. M., Mathematisches Seminar der Universität

(Eingegangen am 19. Februar 1959)

VIDAV, I.
Math. Zeitschr. 71, 458–460 (1959)

Zur Axiomatik des Hilbertschen Raumes

Von
IVAN VIDAV

Ein Hilbertscher Raum $\mathfrak{H}$ ist bekanntlich ein Banachscher Raum, in dem die Norm dem sog. Parallelogrammgesetz $\|x+y\|^2+\|x-y\|^2=2\|x\|^2+2\|y\|^2$ genügt. Das innere Produkt (x, y) läßt sich dann in bekannter Weise einführen. Wenn man aber umgekehrt in $\mathfrak{H}$ zuerst das innere Produkt definiert, ist es nicht notwendig, $\mathfrak{H}$ als Vektorraum vorauszusetzen. Es genügt, die Existenz der Addition der Vektoren und die Multiplikation der Vektoren mit skalaren Größen zu postulieren. Die Addition soll dabei nur einer ganz schwachen Forderung genügen. Wegen der Existenz des inneren Produktes gibt es in $\mathfrak{H}$ eine genügend große Anzahl von Linearformen. Daraus folgt, daß $\mathfrak{H}$ ein linearer Vektorraum ist. Dies werden wir hier unter sehr allgemeinen Voraussetzungen zeigen.

Sei also $\mathfrak{R}$ ein beliebiger Ring mit Einselement 1 und $\mathfrak{H}$ eine Menge von Elementen, in der zwei Rechenoperationen definiert sind:

I. Eine innere Verknüpfung $\mathfrak{H}\times\mathfrak{H}\to\mathfrak{H}$: Jedes geordnete Paar $x, y\in\mathfrak{H}$ bestimmt eindeutig ein Element $x+y\in\mathfrak{H}$.

II. Eine äußere Verknüpfung $\mathfrak{R}\times\mathfrak{H}\to\mathfrak{H}$: Zu jedem Paar α, x, wobei $\alpha\in\mathfrak{R}$ ist und $x\in\mathfrak{H}$, gehört eindeutig ein Element $\alpha x\in\mathfrak{H}$.

Wir betrachten nun diejenigen Abbildungen $x\to f(x)$ von $\mathfrak{H}$ in $\mathfrak{R}$, die folgenden Bedingungen genügen:

$$f(\alpha x+y)=\alpha f(x)+f(y), \tag{1}$$

$$f(1\cdot x)=f(x). \tag{2}$$

Dabei sind $\alpha\in\mathfrak{R}$ und $x, y\in\mathfrak{H}$ beliebig. Jede solche Abbildung $f(x)$ werden wir eine lineare Form in $\mathfrak{H}$ nennen.

Die triviale Abbildung $f(x)\equiv 0$ ist offenbar linear (hier bedeutet 0 das Nullelement von $\mathfrak{R}$). Dies kann möglicherweise die einzige lineare Form in $\mathfrak{H}$ sein, wenn die Verknüpfungen I und II ganz beliebig sind. Gibt es aber in $\mathfrak{H}$ eine genügende Anzahl von diesen Formen und gestattet die Addition $x+y$ eine Kürzungsregel, dann ist $\mathfrak{H}$ ein $\mathfrak{R}$-Modul. Es gilt nämlich der folgende Satz:

Seien in $\mathfrak{H}$ die Voraussetzungen erfüllt:

(a) *Aus $(-1)\cdot x+x=(-1)\cdot x+y$ folgt $x=y$.*

(b) *Mit etwaiger Ausnahme eines Elementes $z\in\mathfrak{H}$ gibt es für jedes andere Element $x\in\mathfrak{H}$ eine lineare Form f so, daß $f(x)\neq 0$ ist.*

Dann ist $\mathfrak{H}$ ein $\mathfrak{R}$-Modul.

Der Beweis dieses Satzes ist sehr einfach. Sei $x \in \mathfrak{H}$ beliebig. Nach (1) berechnet man

$$f[(-1)\cdot x + x] = -f(x) + f(x) = 0.$$

Da dies für alle Formen f gilt, existiert das Element $\mathfrak{z}$ und es ist $(-1)\cdot x + x = \mathfrak{z}$. Aus (1) und (2) bekommt man noch $(-1)\cdot x + 1\cdot x = \mathfrak{z}$. Es ist daher $(-1)\cdot x + x = (-1)\cdot x + 1\cdot x$. Nach (a) folgt daraus $1\cdot x = x$ für alle $x \in \mathfrak{H}$. Ebenso leicht überzeugt man sich, daß $(-1)\cdot x + (x + \mathfrak{z}) = \mathfrak{z} = (-1)\cdot x + x$ ist, daher $x + \mathfrak{z} = x$. Daraus folgt, daß die Formen $f(x)$ homogen und additiv sind. Es ist nämlich

$$f(\alpha x) = f(\alpha x + \mathfrak{z}) = \alpha f(x) + f(\mathfrak{z}) = \alpha f(x) \tag{3}$$

und

$$f(x + y) = f(1\cdot x + y) = f(x) + f(y). \tag{4}$$

Eine nochmalige Anwendung von (a), (b), (1) und (2) ergibt, daß auch $\mathfrak{z} + x = x$ ist. Ferner hat man $f(0\cdot x) = 0\cdot f(x) = 0$ für alle Formen f. Es ist also $0\cdot x = \mathfrak{z}$. Deshalb werden wir von jetzt ab 0 statt $\mathfrak{z}$ schreiben. Es ist dann $x + 0 = 0 + x = x$ und $0\cdot x = 0$.

Gilt $f(a) = f(b)$ für alle Formen f, so ist $a = b$. Da nämlich $f[(-1)a + b] = -f(a) + f(b) = 0$ ist, hat man $(-1)a + b = 0$. Weil auch $(-1)a + a = 0$ gilt, folgt nach (a), daß $a = b$ ist.

Es ist nun leicht, die üblichen Regeln der Addition und Multiplikation mit Skalaren abzuleiten:

1°. $x + y = y + x$.

2°. $(x + y) + z = x + (y + z)$.

3°. $\alpha(x + y) = \alpha x + \alpha y$.

4°. $(\alpha + \beta)x = \alpha x + \beta x$.

5°. $\alpha(\beta x) = (\alpha\beta)x$.

6°. $1\cdot x = x$ (dies ist schon oben bewiesen).

Diese Rechenregeln beweist man so: Die linke und rechte Seite irgendeiner von diesen Gleichungen bezeichne man mit a und b. Eine leichte Rechnung zeigt, daß $f(a) = f(b)$ für alle linearen Formen f gilt. Es ist also $a = b$. Damit ist bewiesen, daß $\mathfrak{H}$ ein $\mathfrak{R}$-Modul ist.

Anmerkung. Ohne (2) ist der Satz nicht richtig. Nehmen wir nämlich für den Ring $\mathfrak{R}$ und für die Menge $\mathfrak{H}$ die reellen Zahlen. Sei dann $x + y$ die Summe in gewöhnlichem Sinne. Die äußere Verknüpfung $r\cdot x$ soll aber mit der Gleichung $r\cdot x = rx + r$ definiert werden. Die Funktion $f(x) = x + 1$, die nur für $\mathfrak{z} = -1$ verschwindet, genügt der Bedingung (1), jedoch nicht der Bedingung (2). Es ist $1\cdot x = x + 1 \neq x$, also ist $\mathfrak{H}$ kein $\mathfrak{R}$-Modul.

Sei nun $\mathfrak{R}$ der Körper der komplexen Zahlen und $\mathfrak{H}$ eine Menge von Elementen, in der die Verknüpfungen I und II definiert sind. Ferner genüge

die Addition der Forderung (a) des Satzes. Es gebe auf $\mathfrak{H}$ eine Semibilinearform (x, y) mit folgenden Eigenschaften:

1. $(\alpha x + y, z) = \alpha(x, z) + (y, z)$.
2. $(1 \cdot x, y) = (x, y)$.
3. $(y, x) = \overline{(x, y)}$.
4. Es gibt in $\mathfrak{H}$ höchstens ein Element $\mathfrak{z}$, für das $(\mathfrak{z}, \mathfrak{z}) = 0$ ist.

Setzen wir $f_a(x) = (x, a)$, so ist offenbar $f_a(x)$ eine Linearform. Da $f_a(a) = (a, a) \neq 0$ für $a \neq \mathfrak{z}$ gilt, ist die Bedingung (b) des Satzes erfüllt. Daraus folgt, daß $\mathfrak{H}$ ein linearer Vektorraum ist.

Die Semibilinearform ist definit: Aus $(x, x) = \overline{(x, x)}$ sieht man zunächst, daß (x, x) immer reell ist. Nehmen wir nun an, es gebe zwei Elemente a und b mit $(a, a) > 0$ und $(b, b) < 0$. Wir betrachten für reelles ξ die Gleichung

$$(\xi a - b, \xi a - b) = 0,$$

also

$$(a, a)\,\xi^2 - [(a, b) + (b, a)]\,\xi + (b, b) = 0.$$

Diese quadratische Gleichung mit reellen Koeffizienten besitzt zwei verschiedene reelle Wurzeln ξ_1, ξ_2, da $(a, a) \cdot (b, b) < 0$ ist. Es ist also $\xi_1 a - b = 0$ und $\xi_2 a - b = 0$; daraus folgt $(\xi_1 - \xi_2)\, a = 0$, also $a = 0$. Dies aber widerspricht der Voraussetzung $(a, a) > 0$.

Es gilt also entweder immer $(x, x) \geqq 0$ oder immer $(x, x) \leqq 0$, d. h. die Form (x, y) ist definit. Im zweiten Fall, wenn $(x, x) \leqq 0$ ist, nehmen wir $-(x, y)$ statt (x, y), um ein inneres Produkt mit den gewohnten Eigenschaften zu bekommen.

Für die Anregung zur jetzigen allgemeinen Fassung des Satzes bin ich den Herrn Referenten der Zeitschrift zu Dank verpflichtet.

Mathematisches Institut der Universität Ljubljana (Jugoslavien)

(Eingegangen am 24. März 1959)

WIELANDT, H.
Math. Zeitschr. 71, 461–462 (1959)

Zum Satz von Sylow. II

Von
HELMUT WIELANDT

Es sei $\mathfrak{G}$ eine endliche Gruppe, π eine Menge von Primzahlen. Unter einer π-Hall-Gruppe von $\mathfrak{G}$ verstehen wir eine Untergruppe $\mathfrak{G}_\pi$ von $\mathfrak{G}$, in deren Ordnung nur Primzahlen aus π aufgehen, während ihr Index durch keine Primzahl aus π teilbar ist. Wir sagen, daß in $\mathfrak{G}$ der π-Sylow-Satz gilt, wenn in $\mathfrak{G}$ eine π-Hall-Gruppe existiert, die zu jeder π-Untergruppe $\mathfrak{M}$ von $\mathfrak{G}$ eine Konjugierte $\mathfrak{M}^G = G^{-1}\mathfrak{M}G$ enthält ($G \in \mathfrak{G}$). Der π-Sylow-Satz gilt z. B. dann, wenn $\mathfrak{G}$ eine nilpotente π-Hall-Gruppe enthält [6]. Dieses Ergebnis kann verschärft werden:

Satz. *π sei die Vereinigung zweier elementfremder Mengen σ, τ von Primzahlen. $\mathfrak{G}$ enthalte eine nilpotente σ-Hall-Gruppe $\mathfrak{G}_\sigma$ und eine τ-Hall-Gruppe $\mathfrak{G}_\tau$, welche elementweise vertauschbar sind. In $\mathfrak{G}$ gelte der τ-Sylow-Satz. Dann gilt in $\mathfrak{G}$ der π-Sylow-Satz.*

Beweis. Es genügt, den Fall zu behandeln, daß σ aus einer einzigen Primzahl p besteht. Dann ist $\mathfrak{G}_p \times \mathfrak{G}_\tau = \mathfrak{G}_\pi$ eine π-Hall-Gruppe von $\mathfrak{G}$. Sei $\mathfrak{M}$ eine π-Untergruppe von $\mathfrak{G}$, deren Ordnung also nur durch Primzahlen aus π teilbar ist. Wir suchen ein G in $\mathfrak{G}$ mit $\mathfrak{M}^G \leqq \mathfrak{G}_\pi$ und erledigen zunächst zwei Sonderfälle.

(1) In $\mathfrak{M}$ gebe es eine normale τ-Hall-Gruppe $\mathfrak{M}_\tau$. Da in $\mathfrak{G}$ der τ-Sylow-Satz gilt, gibt es ein Element $T \in \mathfrak{G}$ mit $\mathfrak{M}_\tau^T \leqq \mathfrak{G}_\tau$. Wir können daher ohne Beschränkung der Allgemeinheit annehmen, daß von vornherein $\mathfrak{M}_\tau \leqq \mathfrak{G}_\tau$ ist. Wir wählen in $\mathfrak{M}$ eine p-Sylow-Gruppe $\mathfrak{M}_p$. Nach Voraussetzung (1) ist $\mathfrak{M}_p$ eine p-Untergruppe des Normalisators $\mathfrak{N}$ von $\mathfrak{M}_\tau$ in $\mathfrak{G}$. Wegen $\mathfrak{M}_\tau \leqq \mathfrak{G}_\tau$ und der elementweisen Vertauschbarkeit von $\mathfrak{G}_p$ mit $\mathfrak{G}_\tau$ ist $\mathfrak{G}_p$ eine p-Sylow-Gruppe von $\mathfrak{N}$. Daher gibt es nach SYLOW ein $G \in \mathfrak{N}$ mit $\mathfrak{M}_p^G \leqq \mathfrak{G}_p$; wir erhalten

$$\mathfrak{M}^G = (\mathfrak{M}_\tau \mathfrak{M}_p)^G = \mathfrak{M}_\tau \mathfrak{M}_p^G \leqq \mathfrak{G}_\tau \mathfrak{G}_p = \mathfrak{G}_\pi.$$

(2) In $\mathfrak{M}$ gebe es eine normale p-Sylow-Gruppe $\mathfrak{M}_p$, deren Index in $\mathfrak{M}$ eine Potenz einer Primzahl $r \in \tau$ ist. Dann kann $\mathfrak{M}$ in der Form $\mathfrak{M} = \mathfrak{M}_p \mathfrak{M}_r$ zerlegt werden. Ähnlich wie im Fall (1) können wir nach SYLOW $\mathfrak{M}_p \leqq \mathfrak{G}_p$ annehmen und dann innerhalb des Normalisators von $\mathfrak{M}_p$ die Gruppe $\mathfrak{M}_r$ in eine r-Sylow-Gruppe von $\mathfrak{G}_\tau$ hineintransformieren. Hierdurch wird wieder $\mathfrak{M}$ in $\mathfrak{G}_\pi$ hineintransformiert.

(3) Nun sei $\mathfrak{M}$ beliebig. Wir schließen induktiv. Für jede echte Untergruppe $\mathfrak{L}$ von $\mathfrak{M}$ sei schon bewiesen, daß $\mathfrak{L}$ zu einer Untergruppe von $\mathfrak{G}_p \times \mathfrak{G}_\tau$ konjugiert ist; dann ist in $\mathfrak{L}$ jedes r-Element, $r \in \tau$, mit jedem p-Element vertauschbar.

Wenn $\mathfrak{M}$ die Voraussetzung (1) befriedigt, sind wir fertig. Wenn nicht, gibt es in $\mathfrak{M}$ eine p-Untergruppe $\mathfrak{P}$ und ein Element R von Primzahlpotenzordnung r^a, $r \in \tau$, derart daß R im Normalisator, aber nicht im Zentralisator von $\mathfrak{P}$ liegt [2; 4, Theorem 14. 4. 7]. Hieraus folgt, daß $\mathfrak{P}$ und R nicht eine echte Untergruppe $\mathfrak{L}$ von $\mathfrak{M}$ erzeugen können; sie erzeugen die ganze Gruppe $\mathfrak{M}$. Daher ist $\mathfrak{P}$ eine normale p-Sylow-Gruppe von $\mathfrak{M}$, und ihr Index ist r^a. Also erfüllt $\mathfrak{M}$ die Voraussetzung (2) und wir sind am Ziel.

Zum Schluß soll auf offene Fragen hingewiesen werden. Läßt sich der Satz in ähnlicher Weise verallgemeinern, wie dies mit dem in [6] behandelten Sonderfall $\tau = \emptyset$ inzwischen [1, 3, 5] geschehen ist? Genügt es, anstelle der Nilpotenz von $\mathfrak{G}_\sigma$ die Gültigkeit des σ-Sylow-Satzes in $\mathfrak{G}$ vorauszusetzen?

Literatur

[1] BAER, R.: Verstreute Untergruppen endlicher Gruppen. Archiv d. Math. **9**, 7–17 (1958). — [2] FROBENIUS, G.: Über auflösbare Gruppen. V. Sitzungsber. Akad. Wiss. Berlin **1901**, 1324–1329. — [3] GOL'BERG, P. A.: Zum Satz von Wielandt. [Russisch.] Uspehi mat. Nauk **14**, 153–156 (1959). — [4] HALL, M.: The theory of groups. New York 1959. — [5] HALL, P.: Theorems like Sylow's. Proc. London math. Soc. (3) **6**, 286–304 (1956). — [6] WIELANDT, H.: Zum Satz von Sylow. Math. Z. **60**, 407–408 (1954).

Tübingen, Mathematisches Institut der Universität

(Eingegangen am 29. April 1959)

Math. Zeitschr. 71, 463–465 (1959)

A weak property L for pairs of matrices

By
OLGA TAUSSKY

A pair of $n \times n$ matrices A, B with complex numbers as elements is said to have the L-property if $\lambda A + \mu B$ has as characteristic roots $\lambda\alpha_i + \mu\beta_i$, where α_i are the characteristic roots of A and β_i the characteristic roots of B—taken in a special ordering—and when λ, μ run through all complex numbers. Let one of the matrices, say A, be diagonable (i.e. similar to a diagonal matrix). The following condition on B is then necessary for A and B to have property L (see T. S. MOTZKIN and O. TAUSSKY [1] p. 108–109).

Let t be the number of different characteristic roots of A and assume that all the characteristic roots α_i of A are arranged in sets of equal ones. Let m_i be the multiplicity of α_i and β_i the corresponding characteristic root of B. Let $B^* = P^{-1}BP$ when $A^* = P^{-1}AP$ is in diagonal form. Then

$$B^* = \begin{pmatrix} B_{11} & B_{12} & \dots & B_{1t} \\ B_{21} & B_{22} & \dots & B_{2t} \\ \cdot & \circ & \dots & \cdot \\ B_{t1} & B_{t2} & \dots & B_{tt} \end{pmatrix}$$

where B_{ii} is an $m_i \times m_i$ matrix, $i = 1, \dots, t$, and

$$|xI - B_{ii}| = \prod_{r}(x - \beta_r) \tag{1}$$

where β_r are the m_i characteristic roots corresponding to α_i and where I is the unit matrix.

Let us denote this condition as "Property l", since it is weaker than property L in general.

THEOREM 1. *Property l implies property L for $n = 2$.*

PROOF. For $n = 2$ we have two cases here: firstly, the case of A a scalar matrix, which certainly implies property L; secondly the case where A has two different roots. In this case the two diagonal elements of B^* are its characteristic roots. This implies that B^* is triangular, hence A^* and B^*, and then also A and B have property L.

THEOREM 2. *Property l implies commutativity for normal matrices when the transforming matrix P is taken as unitary.*

PROOF. For normal matrices it is known (see N. WIEGMANN [2], PARKER [3]) that under the condition (1) all $B_{ik} = 0$ in B^* for $i \neq k$. Since all B_{ii} are normal too, each of them can be transformed into diagonal form by a similarity which leaves the corresponding scalar matrix in A^* unchanged.

Hence A and B can be transformed to diagonal form simultaneously, hence not only have property L, but even commute.

That property l plays some role is shown by Theorem 3 which is the main result of this note.

THEOREM 3. *Let A, B be a pair of $n \times n$ matrices with complex numbers as elements. Let A be diagonable. Let α_i be the characteristic roots of A and $\gamma_i(z)$ be the characteristic roots of $A + zB$ where z is a complex variable. If*

$$\left.\frac{d\gamma_i(z)}{dz}\right|_{z=0} = \gamma_i'(0) = \beta_i$$

(the characteristic roots of B) for some ordering, then A and B have property l.

PROOF. Assume A already in diagonal form. This is no restriction. Let α_i have multiplicity m_i. Among the n (not necessarily different) functions $\gamma_j(z)$ exactly m_i will have the property that $\gamma_j(0) = \alpha_i$. All $\gamma_j(z)$ satisfy the determinantal equation

$$|A + Bz - I\gamma(z)| = 0. \tag{2}$$

To obtain $\gamma_j'(z)$ for those j for which $\gamma_j(0) = \alpha_i$ we differentiate equation (2) m_i times. When differentiating the left hand side we have to form the sum of all determinants in which m_i rows of (2)—not necessarily different ones—have been differentiated. Any row which contains our α_i is of the form

$$z b_{r1},\ z b_{r2},\ \ldots,\ \alpha_i + z b_{rr} - \gamma_j(z),\ \ldots,\ z b_{rn}.$$

This whole row vanishes for $z = 0$. Hence only the one determinant in which all these m_i rows have been differentiated comes in question for $z = 0$. All the other rows in this determinant have then not been differentiated. For simplicity let α_i correspond to the first m_i rows. Then for $z = 0$

$$0 = \begin{vmatrix} b_{11} - \gamma_i'(0) & b_{12} & \cdot\ b_{1m_i} & \cdots\cdots & b_{1n} \\ b_{21} & b_{22} - \gamma_i'(0) & \cdot\ \cdot & \cdots\cdots & b_{2n} \\ \cdot & \cdot & \cdot\ \cdot & \cdots\cdots & \\ b_{m_i 1} & b_{m_i 2} & \cdot\ b_{m_i m_i} - \gamma_i'(0) & \cdots\cdots & b_{m_i n} \\ 0 & 0 & \cdot\ 0 & \alpha_{m_i+1} - \alpha_i & 0 \\ \cdot & \cdot & \cdot\ \cdot & 0 \ldots\ldots & 0 \\ 0 & 0 & \cdot\ 0 & 0 \ldots\ldots & \alpha_n - \alpha_i \end{vmatrix}$$

This shows that m_i of the $\gamma_i'(0)$ are characteristic roots of the first $m_i \times m_i$ principal minor of B and similar facts hold for the other α_i. On the other hand, each $\gamma_i'(0)$ is equal to some of the β_i by assumption. This proves Theorem 3. The converse is true too provided the derivatives exist [1]).

THEOREM 4. *For $n = 2$ any two matrices A, B for which $\gamma_i'(0)$ (as defined in Theorem 3) is equal to β_i for some ordering have property L.*

[1]) This shows that property l is an infinitesimal property L at $z = 0$.

PROOF. If A is diagonable then this follows immediately from Theorems 1 and 3. There only remains the case where A is similar to a matrix $\begin{pmatrix} \alpha & 1 \\ 0 & \alpha \end{pmatrix}$. In this case we may restrict ourselves to the case where A is already in this form. The matrix $A+Bz$ is then

$$\begin{pmatrix} \alpha + b_{11}z & 1 + b_{12}z \\ b_{21}z & \alpha + b_{22}z \end{pmatrix}$$

Its eigenvalues are

$$\gamma_i(z) = \alpha + z\,\frac{b_{11}+b_{22}}{2} \pm \sqrt{z^2(b_{11}-b_{22})^2 + 4b_{21}z(1+b_{12}z)}/2.$$

For $b_{21}=0$ the matrices A and B have already property L. For $b_{21}\neq 0$ we have $\gamma_i'(0)=\infty$, hence this case does not come under our assumption.

THEOREM 5. *Two normal matrices A and B for which $\gamma_i'(0)$ (as defined in Theorem 3) is equal to β_i in some ordering commute.*

PROOF. This follows immediately from Theorems 2 and 3 since every normal matrix is unitarily similar to a diagonal matrix.

REMARK. A special case of normal matrices are hermitian matrices. For hermitian matrices it is known (RELLICH [4]) that the eigenvalues of $A+Bz$ can be expanded into a power series for a neighborhood of the origin. Hence

$$\gamma_i(z) = \alpha_i + c_{i1}z + c_{i2}z^2 + \cdots.$$

Theorem 5 shows that $c_{i1}=\beta_i$ implies that $c_{i2}=c_{i3}=\cdots=0$.

References

[1] MOTZKIN, T. S., and O. TAUSSKY: Pairs of matrices with property L. Trans. Amer. Math. Soc. **73**, 108–114 (1952). — [2] WIEGMANN, N.: Normal matrices with property L. Proc. Amer. Math. Soc. **4**, 35–36 (1953). — [3] PARKER, W. V.: A note on normal matrices. Amer. Math. Monthly **61**, 330–331 (1954). — [4] RELLICH, F.: Störungstheorie der Spektralzerlegung. I. Math. Annalen **113**, 600–619 (1936). — [5] WIELANDT, H.: Pairs of normal matrices with property L. J. Res. Nat. Bur. Standards **51**, 89–90 (1953).

Dept. of Math., California Institute of Technology, Pasadena, Cal. (U.S.A.)

(Eingegangen am 18. März 1959)

GASCHÜTZ, W.
Math. Zeitschr. 71, 466–468 (1959)

Fixkörper von p-Automorphismengruppen rein-transzendenter Körpererweiterungen von p-Charakteristik

KARL HEINRICH WEISE zum 50. Geburtstag

Von

WOLFGANG GASCHÜTZ

Für die Konstruktion von galoisschen Erweiterungskörpern zu vorgegebener Galoisgruppe ist das folgende Problem von großer Wichtigkeit:

(P) *Es sei $K=k(x_1, \ldots, x_n)$ eine rein-transzendente Körpererweiterung von k, $x_1, \ldots, x_n$ seien algebraisch unabhängig über k, und $\mathfrak{G}$ sei eine endliche lineare Substitutionsgruppe in den Unbestimmten $x_1, \ldots, x_n$ mit Koeffizienten aus k. Die hierdurch in K induzierte Automorphismengruppe werde ebenfalls mit $\mathfrak{G}$ bezeichnet. Ist dann der Fixkörper F von $\mathfrak{G}$ ebenfalls rein-transzendent über k oder, damit gleichwertig, wird F über k von n Elementen erzeugt?*

In den wenigen Spezialfällen[1]), die hierauf untersucht worden sind, ist diese Frage bejahend beantwortet worden. Die folgenden Zeilen knüpfen an die Arbeit [3] an. Dort wird für Charakteristik $K=p\neq 0$ eine ziemlich spezielle Klasse von p-Gruppen $\mathfrak{G}$ behandelt. Es soll hier allgemein bewiesen werden:

Satz 1. *In der Bezeichnungsweise von* **(P)** *sei Charakteristik $K=p\neq 0$ und $\mathfrak{G}$ eine endliche p-Gruppe. Dann wird F über k von n Elementen erzeugt.*

Der Satz folgt aus dem allgemeineren

Satz 2. *Es sei*

$$k=K_0\subseteq K_1\subseteq\cdots\subseteq K_n=K$$

eine Kette von Körpern der Charakteristik $p\neq 0$,

$$K_i=K_{i-1}(u_i), \qquad i=1,\ldots,n$$

und $\mathfrak{G}$ eine endliche p-Automorphismengruppe von K mit den Eigenschaften

(1) $$\varkappa^G=\varkappa \text{ für alle } \varkappa\in k \text{ und } G\in\mathfrak{G}$$

(2_i) $$u_i^G-u_i\in K_{i-1} \quad , i=1,\ldots,n.$$

Dann wird der Fixkörper F von $\mathfrak{G}$ über k von n Elementen erzeugt:

$$F=k(z_1,\ldots,z_n).$$

Beweis. Der Satz gilt, wenn $\mathfrak{G}=1$ oder $n=0$ ist. Er sei bewiesen für alle Gruppenordnungen $<\mathfrak{G}:1$ oder für Kettenlängen $<n$.

[1]) Siehe Literaturverzeichnis am Schluß der Arbeit [1, 2, 3, 4, 5].

(1) und $(2_1), \ldots, (2_{n-1})$ zufolge erfüllt $\mathfrak{G}$ auf K_{n-1} die Induktionsannahmen. Der zugehörige Fixkörper F_{n-1} wird also über k von $n-1$ Elementen erzeugt; etwa $F_{n-1} = k(z_1, \ldots, z_{n-1})$.

$\mathfrak{H}$ sei der Normalteiler der Elemente von $\mathfrak{G}$, die K_{n-1} elementweise festlassen.

Fall 1: $\mathfrak{H} = 1$. Dann existiert nach einem Satz von E. WITT [*1*, I(2)] ein Element $y_{n-1} \varepsilon K_{n-1}$ mit

$$u_n^G - u_n = y_{n-1}^G - y_{n-1}$$

für alle $G \varepsilon \mathfrak{G}$. Unter $\mathfrak{G}$ ist $z_n = u_n - y_{n-1}$ invariant und daher $F_{n-1}(z_n)$ elementweise invariant. Nun ist

$$K_{n-1}(z_n) : F_{n-1}(z_n) \leqq K_{n-1} : F_{n-1}$$

und $K_{n-1}(z_n) = K$. Nach dem Hauptsatz der Galoistheorie ist

$$\mathfrak{G} : 1 \leqq K : F_{n-1}(z_n)$$

und

$$K_{n-1} : F_{n-1} = \mathfrak{G} : 1;$$

also ist $F_{n-1}(z_n) = k(z_1, \ldots, z_{n-1}, z_n)$ der Fixkörper von $\mathfrak{G}$ in K.

Fall 2: $\mathfrak{H} \neq 1$. Dann sei $\mathfrak{M}$ ein minimaler Normalteiler von $\mathfrak{G}$ in $\mathfrak{H}$. $\mathfrak{M}$ liegt im Zentrum von $\mathfrak{G}$, hat die Ordnung p und ist zyklisch: $\mathfrak{M} = \langle M \rangle$.

Es werde

$$u_n^G - u_n = c_G, \qquad c_G \varepsilon K_{n-1},$$

gesetzt. Aus $u_n^{GM} = u_n^{MG}$ für alle $G \varepsilon \mathfrak{G}$ folgt:

$$u_n + c_M + c_G^M = (u_n + c_G)^M = u_n^{GM} = u_n^{MG} = (u_n + c_M)^G = u_n + c_G + c_M^G;$$

da K_{n-1} unter $\mathfrak{M}$ elementweise invariant bleibt, ist also

$$c_M^G = c_M \tag{3}$$

für alle $G \varepsilon \mathfrak{G}$. Ferner gilt für $v_n = u_n / c_M$:

$$v_n^M = v_n + 1.$$

Ist K^* der Fixkörper von $\mathfrak{M}$, so ist also

$$u_n^* = v_n^p - v_n \varepsilon K_n^*$$

und

$$K : K^* \leqq K : K_{n-1}(u_n^*) \leqq p.$$

Nach der Galoistheorie ist aber $K : K^* = p$. Das ergibt $K^* = K_{n-1}(u_n^*)$. In K^* wird nun durch $\mathfrak{G}$ eine zu $\mathfrak{G}/\mathfrak{M}$ isomorphe Gruppe $\mathfrak{G}^*$ von Automorphismen induziert, die neben (1), $(2_1), \ldots, (2_{n-1})$ auch die Bedingung (2_n) für u_n^* an Stelle von u_n erfüllt:

$$u_n^{*G} = \left(\frac{u_n + c_G}{c_M^G}\right)^p - \frac{u_n + c_G}{c_M^G};$$

auf Grund von (3) und Charakteristik $K=p$ ist

$$u_n^{*G} = u_n^* + c_G^*, \qquad c_G^* \varepsilon K_{n-1}.$$

Da F auch der Fixkörper von K^* unter $\mathfrak{G}^*$ und $\mathfrak{G}^*:1 < \mathfrak{G}:1$ ist, folgt nun der Satz aus den Induktionsvoraussetzungen; w.z.b.w.

Satz 1 ergibt sich leicht aus Satz 2 mit der bekannten Tatsache, daß eine endliche lineare p-Substitutionsgruppe über einem Körper der Charakteristik p durch eine nichtsinguläre lineare Transformation der Unbestimmten x_i in gewisse u_i in den neuen Unbestimmten die Form

$$u_i^G = u_i + \sum_{k=1}^{i-1} \gamma_{ik} u_k, \qquad \gamma_{ik} \varepsilon k, \qquad i = 1, \ldots, n$$

hat.

Literatur

[1] BURNSIDE, W.: Theory of groups of finite Order, Chapt. XVII. Cambridge 1911. — [2] KUNIYOSHI, H.: On purely transcendency of a certain field. Tôhoku Math. J. **6**, 101–108 (1955). — [3] KUNIYOSHI, H.: Certain subfields of rational function fields. Proc. Intern. Symp. on alg. Number Theory, 241–243, Tokyo-Nikko 1955. — [4] MASUDA, K.: On a problem of Chevalley. Nagoya Math. J. **8**, 59–63 (1955). — [5] MORIKAWA, H.: On the invariants of finite nilpotent groups. Osaka Math. J. **10**, 53–56. — [6] WITT, E.: Konstruktion von galoisschen Körpern der Charakteristik p zu vorgegebener Gruppe der Ordnung p^l. J. reine u. angew. Math. **174**, 237–245 (1935).

Kiel, Mathematisches Seminar der Universität

(Eingegangen am 18. Mai 1959)